DUKE MATHEMATICAL JOURNAL

Volume 5, Number 4
DECEMBER, 1939

DUKE UNIVERSITY PRESS
DURHAM, N. C.

DUKE MATHEMATICAL JOURNAL

This periodical is published quarterly under the auspices of Duke University by Duke University Press at Durham, North Carolina. It is printed at Mt. Royal and Guilford Avenues, Baltimore, Maryland, by the Waverly Press.

Entered as second class matter at the Post Office, Durham, North Carolina. Additional entry at the Post Office, Baltimore, Maryland.

The subscription price for the current volume is four dollars, postpaid; back volumes, five dollars each, carriage extra; single numbers, one dollar and twenty-five cents, postpaid. Subscriptions, orders for back numbers, and notice of change of address should be sent to Duke University Press, Durham, North Carolina.

Individual and institutional members of the Mathematical Association of America may subscribe to the current volume at half price. During 1940, but not thereafter, they are also entitled to complete their back files at the same reduction. To get the reduced price, orders for subscriptions and back numbers must bear the mention "Member MAA." If an order at the reduced price is placed through an agent, the purchaser must pay any commission charge incurred.

Manuscripts and editorial correspondence should be addressed to Duke Mathematical Journal, 4785 Duke Station, Durham, North Carolina. An Author's Manual containing detailed information about the preparation of papers for publication will be sent on request.

Authors are entitled to one hundred free reprints. Additional copies will be supplied at cost. All reprints will be furnished with covers unless the contrary is specifically requested.

The American Mathematical Society is officially represented on the Editorial Board by Professors Ore and Whyburn.

Made in United States of America

WAVERLY PRESS, INC.
BALTIMORE, U. S. A.

THE ALGEBRAIC THEORY OF DIABOLIC MAGIC SQUARES

By Barkley Rosser and R. J. Walker

1. **Introduction.** For many centuries magic squares have attracted the attention of people interested in mathematics. However, no general theory of magic squares has ever been developed, principally because such a theory would have to involve partition problems of a peculiarly difficult character. If one relaxes the condition that the elements of the square shall be the first n^2 positive integers, the study of magic squares becomes a purely algebraical problem involving linear equations in n^2 variables. For diabolic magic squares, in which all diagonals have the same sum, the group of transformations which leave the algebraic conditions unchanged is much larger than it is for ordinary magic squares, in which only the two main diagonals have this sum, and the resulting theory is more interesting. Hence we have confined our attention to diabolic squares and their generalizations.

In the first part of the paper the general algebraic theory of diabolic squares is considered. A group of transformations which carries any diabolic square of order n into a diabolic square is constructed; for n a prime $\geqq 7$ it is shown that this is the largest group carrying the most general diabolic square into a diabolic square. The latter result is proved by obtaining, for a square of prime order, the general solution of the $4n$ linear equations which the elements of the square must satisfy.

The latter part of the paper is concerned with the applications of these results to squares whose elements are the integers $1, \cdots, n^2$. The principal result here is that there are precisely 28,800 diabolic squares of order 5, all of an easily constructed type which we have called regular. The existence or non-existence of regular and irregular diabolic squares is demonstrated for squares of all orders.

2. **Some definitions.** By a *square* of order n, S_n, we shall mean a square matrix of order n whose elements are members of any field K of characteristic prime to n. The elements of S_n shall be denoted by A_{ij} or by $A(i, j)$, i denoting the row and j the column, and whenever i or j is not in the range 1 to n inclusive, we shall understand that it is to be reduced to its least positive residue modulo n. The square whose elements are A_{ij} shall often be designated by A.

A *configuration* in S_n is any linear combination of elements

$$\sum_{x=1}^{r} \alpha_x A(i_x, j_x),$$

the α_x being members of K. S_n is said to *admit* the configuration if

$$\sum_{x=1}^{r} \alpha_x A(i + i_x, j + j_x) = N \sum \alpha_x$$

Received November 2, 1938; in revised form, September 25, 1939.

for all i and j, N being independent of i and j. By summing this equation over i and j, we see that necessarily

$$N = \frac{1}{n^2} \sum_{i,j=1}^{n} A_{ij}$$

if $\sum \alpha_z \neq 0$. If $\sum \alpha_z = 0$, N is not determined and may be given the above value. If S_n admits a certain set of configurations, so does the square with elements

$$A'_{ij} = A_{ij} - \frac{1}{n^2} \sum_{i,j=1}^{n} A_{ij},$$

and for this square $N = 0$. Hence, when we are dealing with a square for which we have assumed merely that it admits certain configurations, it will involve no loss of generality to assume $N = 0$.

The configurations with which we are especially concerned are the sums of n elements,

$$\sum_{x=1}^{n} A(i + ax, j + bx),$$

where $(a, b, n) = 1$. (We shall use the standard notation $(x, y, \cdots, z)$ for the highest common factor of $x, y, \cdots, z$.) These are denoted by $P(i, j; a, b)$ and are called *paths*. We are often not interested in a particular path $P(i, j; a, b)$ but rather in the whole set of n paths with the given a, b. In this case we shall speak of "the path $P(a, b)$", meaning thereby any one of this set of paths.

If $(c, n) = 1$, the paths $P(i, j; a, b)$ and $P(i, j; ca, cb)$ contain the same elements, so that $P(ca, cb) = P(a, b)$.

The most important paths are $P(0, 1)$, $P(1, 0)$, $P(1, 1)$, and $P(1, -1)$. These are respectively the rows, columns, and two sets of diagonals of S_n. If S_n admits these four paths, it is said to be a *diabolic square* (d.s.).

An interesting set of configurations is

$$A_{ij} + A_{kl} - A_{il} - A_{kj}.$$

A square which admits all these configurations is called a *primitive square* (p.s.).

If d is a factor of n, the configuration

$$\sum_{x,y=1}^{n/d} A(i + dx, j + dy)$$

is called a *lattice* of order n/d, and is denoted by $L(i, j; d)$. When the particular i and j do not matter, the lattice is denoted by $L(d)$. We shall also use the word "lattice" and the symbol $L(i, j; d)$ to refer to the square of order n/d whose elements are $B(x, y) = A(i + dx, j + dy)$.

THEOREM 2.1. *A p.s. of order n admits all paths $P(a, b)$ with $(ab, n) = 1$.*

This follows readily from the definition of a p.s.

THEOREM 2.2. *Each lattice of a p.s. is a p.s.*

Proof.

$$A(i + dx, j + dy) + A(i + ds, j + dt) - A(i + dx, j + dt) - A(i + ds, j + dy) = 0.$$

THEOREM 2.3. *Let a square of order mn admit the paths*

$$P(a_i , b_i) \qquad (i = 1, 2, \cdots , s).$$

If a square of order n which admits these same paths necessarily admits the configuration

$$\sum_{z=1}^{t} \alpha_z A(i_z , j_z),$$

then the square of order mn admits the configuration

$$\sum_{z=1}^{t} \alpha_z L(i_z , j_z ; n).$$

Proof. Let A be a square of order mn admitting the paths

$$P(a_i , b_i) \qquad (i = 1, 2, \cdots , s).$$

Suppose $N = 0$. Define a square B, of order n, by

$$B(i, j) = L(i, j; n).$$

To prove our theorem, it is clearly sufficient to prove that B admits the paths $P(a_i , b_i)$. Let $P_i(u, v)$ denote the path $P(u, v; a_i , b_i)$ taken from B. Then

$$P_i(u, v) = \sum_{z=1}^{n} \sum_{x, y=1}^{m} A(u + a_i z + nx , v + b_i z + ny).$$

Since $P(a_i , b_i)$ is a path in A, $(a_i , b_i , mn) = 1$. Choose α_i and β_i so that $a_i\beta_i - \alpha_i b_i \equiv 1 \pmod{mn}$. Let $Q_i(u, v)$ denote

$$\sum_{k=1}^{m} P(u + k\alpha_i n, v + k\beta_i n; a_i , b_i)$$

taken from A. It is easily shown that there are exactly m^2n elements in $Q_i(u, v)$, and that every element of $Q_i(u, v)$ is an element of $P_i(u, v)$. As there are at most m^2n elements of $P_i(u, v)$, $P_i(u, v)$ and $Q_i(u, v)$ are the same. So $P_i(u, v) = 0$.

THEOREM 2.4. *A d.s. of even order admits* $L(2)$.

Proof. In a d.s. of order 2 with $N = 0$,

$$2A_{11} = P(1, 1; 1, 1) + P(1, 1; 0, 1) - P(1, 2; 1, 0) = 0.$$

So a d.s. of order 2 admits A_{ij}. So, by Theorem 2.3, a d.s. of order $2m$ admits $L(i, j; 2)$.

THEOREM 2.5. *A d.s. of order 3m admits $L(3)$.*

Proof. In a d.s. of order 3 with $N = 0$,

$$3A_{11} = P(1, 1; 1, 1) + P(1, 1; 1, -1) + P(1, 1; 0, 1) - P(1, 2; 1, 0) - P(1, 3; 1, 0) = 0.$$

So a d.s. of order 3 admits A_{ij}. The theorem follows by Theorem 2.3.

THEOREM 2.6. *A d.s. of order 4m admits $L(i, j; 4) + L(i + 2, j + 2; 4)$.*

Proof. A d.s. of order 4 admits $A(i, j) + A(i + 2, j + 2)$.[1] The theorem follows from Theorem 2.3.

THEOREM 2.7. *Suppose that, for a given d, each lattice $L(i, j; d)$ of a square A is diabolic, the value of a path in $L(i, j; d)$ being N_{ij}. If the square of order d with N_{ij} in the i-th row and j-th column is diabolic, then A is diabolic.*

Proof. Consider, for example, $P(1, -1)$.

$$\begin{aligned} P(i, j; 1, -1) &= \sum_{x=1}^{n} A(i + x, j - x) \\ &= \sum_{y=1}^{d} \sum_{z=1}^{n/d} A(i + y + dz, j - y - dz) \\ &= \sum_{y=1}^{d} N_{i+y, j-y}. \end{aligned}$$

Hence $P(i, j; 1, -1)$ is independent of i and j.

3. Transformations.

Let

$$T = \begin{Vmatrix} a & c \\ b & d \end{Vmatrix}$$

be a matrix of integers with $|T|$ prime to n. The transform of the square A by T is defined to be the square B, where $B(ai + cj, bi + dj) = A(i, j)$. Clearly this just amounts to transforming each vector (i, j) into the vector $(ai + cj, bi + dj)$ by means of T, and so has the usual characteristics of a transformation of vectors by matrices; in particular,

(a) the result of transforming first by T and then by S is the same as the result of transforming by the matrix product ST;

(b) since $|T|$ is prime to n, one can find an ϵ such that $\epsilon(ad - bc) \equiv 1 \pmod{n}$, and so

$$T^{-1} = \begin{Vmatrix} d\epsilon & -c\epsilon \\ -b\epsilon & a\epsilon \end{Vmatrix};$$

[1] Barkley Rosser and R. J. Walker, *On the transformation group for diabolic magic squares of order four*, Bull. Am. Math. Soc., vol. 44(1938), pp. 416–420. See Theorem 2.

(c) $B(ai + cj + (ae + cf)x, bi + dj + (be + df)x) = A(i + ex, j + fx)$, so that T transforms the path $P(e, f)$ into the path $P(ae + cf, be + df)$;

(d) $B(ai + cj + (ax + cy)f, bi + dj + (bx + dy)f) = A(i + fx, j + fy)$, so that T transforms the lattice $L(f)$ into a lattice $L(f)$.

THEOREM 3.1. *The transform of a p.s. of order n by*

$$T = \begin{Vmatrix} a & c \\ b & d \end{Vmatrix}$$

is diabolic if $abcd(a^2 - b^2)(c^2 - d^2)$ is prime to n.

Proof. The paths which transform into $P(0, 1)$, $P(1, 0)$, $P(1, 1)$, $P(1, -1)$ are obtained by transforming these by T^{-1}. Removing the factor ϵ, since $(\epsilon, n) = 1$, we see that these are $P(-c, a)$, $P(d, -b)$, $P(-c + d, a - b)$, $P(c + d, -a - b)$. The theorem follows from Theorem 2.1.

A transformation satisfying the conditions of Theorem 3.1 shall be called *regular*.

From its definition, a p.s. remains primitive under any permutation of rows and columns among themselves. Two p.s. obtainable from one another by such a permutation are said to be of the same type. There are therefore $(n!)^2$ p.s. of order n of any given type if the elements of the square are all different.

We wish to determine the number of d.s. obtainable by regular transformations from p.s. of a given type whose elements are all different. We first note that two different p.s. of the same type may give the same d.s. when acted on by two different regular transformations. For if A is a p.s., T a regular transformation, and

$$S = \begin{Vmatrix} \alpha & 0 \\ 0 & \beta \end{Vmatrix} \qquad ((\alpha\beta, n) = 1),$$

then SA is primitive and of the same type as A, TS is regular, and $(TS)A = T(SA)$. Conversely, if A and B are p.s. of the same type and $T_1A = T_2B$, then $A = T_1^{-1}T_2B$, so that $T_1^{-1}T_2$ transforms rows into rows and columns into columns. But the only matrix transformations which do this are those of the form of S. Hence $T_1^{-1}T_2 = S$, or $T_2 = T_1S$. Now if T is regular, we can find β, δ such that $b\beta \equiv d\delta \equiv 1 \pmod{n}$, and then

$$T' = T \begin{Vmatrix} \beta & 0 \\ 0 & \delta \end{Vmatrix} = \begin{Vmatrix} a\beta & c\delta \\ 1 & 1 \end{Vmatrix}.$$

A regular transformation for which $b \equiv d \equiv 1 \pmod{n}$ shall be called *normal*. Since obviously no two normal transformations can be related by $T_1 = T_2S$ unless S is the identity matrix, the number of d.s. which are regular transforms of p.s. of a given type with elements all different is $(n!)^2\theta(n)$, where $\theta(n)$ is the number of normal transformations of order n. We wish to determine the function $\theta(n)$.

The following lemma is readily proved.

LEMMA 3.1. *Let $F(x_1, \cdots, x_r)$ be any function of integers assuming integer values such that for all n*

$$F(x_1, \cdots, x_r) \equiv F(y_1, \cdots, y_r) \pmod{n}$$

if $x_i \equiv y_i \pmod{n}$. If $\theta(n)$ denotes the number of sets $(a_1, \cdots, a_r)$ $(0 \leqq a_i \leqq n - 1)$ such that $F(a_1, \cdots, a_r)$ is prime to n, then

(a) $\theta(mn) = \theta(m)\theta(n)$ *if* $(m, n) = 1$;

(b) $\theta(p^s) = p^{r(s-1)}\theta(p)$ *if p is prime.*

THEOREM 3.2. *There are $(n!)^2\theta(n)$ d.s. which are regular transforms of p.s. of a given type whose elements are all different, where*

(a) $\theta(mn) = \theta(m)\theta(n)$ *if* $(m, n) = 1$,

(b) $\theta(p^s) = p^{2s-2}(p - 3)(p - 4)$ *if p is an odd prime,*

(c) $\theta(2^s) = 0$.

Proof. Using the definition of a regular transformation and the fact that the determinant of the transformation must be prime to n, we see that the matrix

$$\begin{Vmatrix} a & c \\ 1 & 1 \end{Vmatrix}$$

defines a regular transformation if and only if

$$F(a, c) = ac(a^2 - 1)(c^2 - 1)(a - c)$$

is prime to n. Any polynomial with integer coefficients satisfies the condition of Lemma 3.1, and so we have only to determine the value of $\theta(p)$ for the function $F(a, c)$. a can have any of the values $2, 3, \cdots, p - 2$, and for each choice of a, c can have any of these values except that assumed by a. Hence $\theta(2) = 0$ and $\theta(p) = (p - 3)(p - 4)$ if $p > 2$, and the theorem follows.

It is well known that any d.s. is transformed into a d.s. by rotation through a multiple of 90°, reflection in a diagonal, or a cyclic permutation of the rows and columns. The following theorem gives a more complete view of the situation.

THEOREM 3.3. *The permutations of the elements of a square of order $n \geqq 4$ which take rows, columns, and the two sets of diagonals (r, c, d, d') into r, c, d, d', not necessarily in the same order, form a group G which is generated by*

$$L = \begin{cases} i' = i - 1 \\ j' = j \end{cases}, \qquad M = \begin{cases} i' = i \\ j' = j - 1 \end{cases}, \qquad S_\alpha = \begin{Vmatrix} \alpha & 0 \\ 0 & \alpha \end{Vmatrix} \qquad ((\alpha, n) = 1),$$

$$O = \begin{Vmatrix} 1 & 0 \\ 0 & -1 \end{Vmatrix}, \qquad P = \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix}, \qquad Q = \begin{Vmatrix} 1 & 1 \\ 1 & -1 \end{Vmatrix},$$

if n is odd, and by L, M, S_α, O, P if n is even.

Proof. Each of these permutations is easily seen to transform r, c, d, d' into r, c, d, d'. We have to show, conversely, that if a permutation T has this property, it is a member of G.

First of all, if T transforms one row into a row, it transforms all rows simi-

larly; for two rows have no common element and hence their transforms cannot. The same is true of columns or diagonals. Let us consider first the case where T transforms r, c, d into r, c, d respectively, nothing being said about what d' goes into. Let

$$T = \begin{cases} i' = f(i, j) \\ j' = g(i, j) \end{cases}.$$

Since $Tr = r$, i' must be independent of j, and since $Tc = c$, j' must be independent of i. Hence

$$T = \begin{cases} i' = f(i) \\ j' = g(j) \end{cases}.$$

If $f(0) = a$, $g(0) = b$ are not both zero, consider the transformation $T' = L^a M^b T$, which also takes r, c, d into r, c, d respectively. Evidently

$$T' = \begin{cases} i' = f'(i) \\ j' = g'(j) \end{cases}$$

and $f'(0) = g'(0) = 0$. Since $T'd = d$, the transform by T' of the d

$$A(i, 0), \quad A(i + 1, 1), \quad \cdots, \quad A(i + j, j), \quad \cdots$$

is a d. This requires that

$$f'(i + j) - g'(j) = f'(i) - g'(0) = f'(i).$$

Put $i = 0$; then $f'(j) = g'(j)$, and we have

$$f'(i + j) = f'(i) + f'(j),$$

which relation implies $f'(i) = \alpha i$, and so $L^a M^b T = S_\alpha$, or $T = M^{-b} L^{-a} S_\alpha$. Since T' is a permutation it has an inverse, and this evidently requires that $(\alpha, n) = 1$.

Now let T take r, c, d, d' into r, c, d, d' in any order whatever. If $Tr = c$, then $(PT)r = r$. If n is even, $Tr \neq d$ or d', for a d and a d' have two or no common elements while an r has only one element in common with any c, d, or d'. If n is odd and $Tr = d'$, then $(QT)r = r$; if $Tr = d$, $(PQT)r = r$. Hence in any case there is an element H of G such that if $T'' = HT$, $T''r = r$. If $T''c = c$ and $T''d = d'$, then $(OT'')r = r$, $(OT'')c = c$, and $(OT'')d = d$, and so OT'', and hence T, belongs to G by the previous paragraph. If $T''c = d$, $T''d = d'$, then n is odd and WT'', where

$$W = \begin{Vmatrix} 2 & 0 \\ 1 & -1 \end{Vmatrix},$$

takes r, c, d into r, c, d, respectively, and hence is an element of G; this is impossible, for since $T''d' = c$ we would have $(WT'')d' = Wc = P(2, 1)$. If $T''c = d$, $T''d = c$, we have the same result, using $WT''O$. If $T''c = d'$ we treat OT'' in the same manner. This exhausts all the possibilities, and in every case T is a member of G.

THEOREM 3.4. *The group G is of order $4n^2\varphi(n)$ if n is even and $8n^2\varphi(n)$ if n is odd, $\varphi(n)$ being the number of integers less than n and prime to n.*

Proof. Suppose n is odd. We shall show that every element of G can be expressed in the unique form

$$L^aM^bS_\alpha O^eP^fQ^g,$$

where $0 \leqq a \leqq n-1; 0 \leqq b \leqq n-1; 0 < \alpha < n; e = 0, 1; f = 0, 1; g = 0, 1$. The following relations are easily verified:

(a) $$ML = LM,$$

(b) $$\begin{Vmatrix} a & c \\ b & d \end{Vmatrix} L = L^aM^b \begin{Vmatrix} a & c \\ b & d \end{Vmatrix},$$
$$\begin{Vmatrix} a & c \\ b & d \end{Vmatrix} M = L^cM^d \begin{Vmatrix} a & c \\ b & d \end{Vmatrix},$$

(c) $$S_\alpha S_\beta = S_{\alpha\beta},$$

(d) $$\begin{Vmatrix} a & c \\ b & d \end{Vmatrix} S_\alpha = S_\alpha \begin{Vmatrix} a & c \\ b & d \end{Vmatrix},$$

(e) $$PO = S_{n-1}OP,$$

(f) $$L^n = M^n = O^2 = P^2 = S_1,$$

(g) $$Q^2 = S_2,$$

(h) $$QO = PQ, \qquad QP = OQ.$$

Now given any product of powers of the generators we can transfer all the Q's to the right end by the use of (b), (d), and (h), and use (g) to reduce the exponent of Q to 0 or 1. Next we use (b), (d), and (e) to bring the P's next to the Q's, and by (f) we reduce the exponent of P to either 0 or 1. Then the O's are brought next to the P's, and again the exponent is reduced to 0 or 1. Finally the S's are brought together and combined, and L, M put in the proper order. L and M are both of order n, and there are $\varphi(n)$ possibilities for α, so the total possible number of such normal forms is $8n^2\varphi(n)$. We have now only to prove that no two such normal forms can represent the same element of the group. Suppose

$$L^aM^bS_\alpha O^eP^fQ^g = L^{a'}M^{b'}S_{\alpha'}O^{e'}P^{f'}Q^{g'}.$$

If $g \neq g'$, Q can be expressed as a product of L, M, S_α, O, P. But these change r into r or c, while Q takes r into d. Hence $g = g'$, and the Q's can be omitted from the equation. If $f \neq f'$, P is a product of L, M, S_α, O, all of which take r into r, whereas P takes r into c. Hence $f = f'$ and the P's can be omitted. If $e \neq e'$, O is a product of L, M, S_α, all of which take d into d, while O takes d into d'. Hence $e = e'$, and the O's can be omitted. $S_{\alpha\beta}$, where $\beta\alpha' \equiv 1$

(mod n), is then a product of L, M both of which preserve the cyclic order of the rows; this requires $\alpha\beta \equiv 1 \pmod{n}$, or $\alpha = \alpha'$. L and M are obviously independent.

If n is even the same argument applies, with the omission of all mention of Q.

THEOREM 3.5. *The set of $(n!)^2\theta(n)$ d.s. of Theorem 3.2 is invariant under G.*

Proof. Let A be a p.s. and R a regular transformation; then RA is diabolic. We wish to show that if T is any member of G, there exists a regular transformation R' and a p.s. A' of the same type as A, such that $TRA = R'A'$. This shall be done by showing that if T is any generator of G, there is a regular R' such that $TR = R'T'$, where T' is a permutation of rows and columns. Let

$$R = \begin{Vmatrix} a & c \\ b & d \end{Vmatrix}.$$

From (a) and (b) of Theorem 3.4 we obtain

$$RL^{\alpha}M^{\beta} = L^{\alpha a+\beta c}M^{\alpha b+\beta d}R.$$

Solving $\alpha a + \beta c \equiv 1$, $\alpha b + \beta d \equiv 0 \pmod{n}$, we get $\alpha \equiv \epsilon d$, $\beta \equiv -\epsilon b$, $\epsilon(ad - bc) \equiv 1 \pmod{n}$, so that

$$LR = RL^{\epsilon d}M^{-\epsilon b}.$$

Similarly,

$$MR = RL^{-\epsilon c}M^{\epsilon a}.$$

By matrix multiplication we obtain

$$S_aR = RS_a, \quad OR = \begin{Vmatrix} a & c \\ -b & -d \end{Vmatrix}, \quad PR = \begin{Vmatrix} b & d \\ a & c \end{Vmatrix}, \quad QR = \begin{Vmatrix} a+b & c+d \\ a-b & c-d \end{Vmatrix}.$$

Each of the new matrices is seen to be regular (the last, because n is odd), and since L, M, and S_a merely permute rows and columns among themselves, the theorem is proved.

4. **Squares of prime order.** In this section we shall consider some special properties of squares whose order is a prime p.

The stipulation that a square S_n admit a given set of configurations implies that the elements of S_n satisfy a certain set of linear equations. The general solution of these equations, the elements being linear functions of a certain number of essential, independent parameters, shall be called the *general square* of order n admitting the given configurations.

LEMMA 4.1. *The general square of prime order p admitting all paths is $A_{ij} = N$.*[2]

[2] This theorem is true for squares of any order. The proof is very complicated, so it is not included here. A typewritten copy of the proof has been deposited in the Cornell University library as part of a monograph bearing the title *Magic Squares, Supplement*, by J. B. Rosser and R. J. Walker. This monograph will be referred to as "the supplement".

Proof.

$$\begin{aligned}
pN &= \sum_{b=1}^{p} P(i, j; 1, b) - \sum_{x=1}^{p-1} P(i + x, 0; 0, 1) \\
&= \sum_{b=1}^{p} \left(\sum_{x=1}^{p} A(i + x, j + bx)\right) - \sum_{x=1}^{p-1} \left(\sum_{y=1}^{p} A(i + x, y)\right) \\
&= \sum_{b=1}^{p} A(i, j) + \sum_{x=1}^{p-1} \left(\sum_{b=1}^{p} A(i + x, j + bx)\right) - \sum_{x=1}^{p-1} \left(\sum_{y=1}^{p} A(i + x, y)\right) \\
&= pA(i, j) + \sum_{x=1}^{p-1} \left(\sum_{k=1}^{p} A(i + x, k)\right) - \sum_{x=1}^{p-1} \left(\sum_{y=1}^{p} A(i + x, y)\right) \\
&= pA(i, j).
\end{aligned}$$

LEMMA 4.2. *The number of essential, independent parameters in the general square S_p of prime order admitting r sets of paths is $p^2 - (p - 1)r$.*

Proof. Let s be the number of independent equations among the pr equations expressing the fact that S_p admits the r sets of paths in question. Then the number of essential, independent parameters in the general solution is $p^2 - s$. Now for S_p to admit the path $P(a, b)$ means that for each i and j

$$\sum_{x=1}^{p} A(i + ax, j + bx) = \frac{1}{p} \sum_{u,v} A(u, v).$$

Hence we see that if S_p admits all but one of the set of paths $P(a, b)$, it automatically admits the remaining one. Hence at least r of the pr equations are dependent, and so $s \leqq (p - 1)r$. Suppose $s < (p - 1)r$. If we adjoin the equations for the remaining paths, the number of independent equations in the resulting set is $p^2 - 1$ by Lemma 4.1. Hence if t is the number of independent added equations, $s + t \geqq p^2 - 1$, and so $t > (p - 1)(p + 1 - r)$. However, there are exactly $p + 1$ sets of paths, namely, $P(0, 1)$, $P(1, 0)$, $P(1, 1)$, $\cdots$, $P(1, p - 1)$. Therefore $p(p + 1 - r)$ equations were added, of which at least $p + 1 - r$ are dependent, so $t \leqq (p - 1)(p + 1 - r)$. From this contradiction $s = (p - 1)r$, and the lemma follows.

LEMMA 4.3. *In a square of prime order p two paths of different sets have exactly one common element.*

Proof. If $P(i, j; a, b)$ and $P(k, l; c, d)$ are in different sets, $ad - bc \not\equiv 0 \pmod p$. Hence the congruences

$$\left.\begin{aligned} xa - yc &\equiv k - i \\ xb - yd &\equiv l - j \end{aligned}\right. \qquad (\text{mod } p)$$

have a unique solution. This implies that the two paths have the single element

$$A(i + ax, j + bx) = A(k + cy, l + dy)$$

in common.

Let $P(a_1, b_1), P(a_2, b_2), \cdots, P(a_{p+1}, b_{p+1})$ be an ordering of the sets of paths of S_p. For each s, choose α_s and β_s so that $\alpha_s b_s - a_s \beta_s \not\equiv 0 \pmod{p}$. Then define A^s for $s = 1, 2, \cdots, p + 1$ to be the square such that for all values of z

$$A^s(za_s + y\alpha_s, zb_s + y\beta_s) = x_{(p-1)(s-1)+y} \qquad (y = 1, 2, \cdots, p - 1),$$

$$A^s(za_s, zb_s) = -\sum_{y=1}^{p-1} x_{(p-1)(s-1)+y}.$$

That is, each path of the set $P(a_s, b_s)$ for A^s has all its elements filled with either

$$x_{(p-1)(s-1)+1}, x_{(p-1)(s-1)+2}, \cdots, x_{(p-1)s}, \text{ or } -\sum_{y=1}^{p-1} x_{(p-1)(s-1)+y}.$$

Also define A^0 by

$$A^0_{ij} = x_0.$$

Clearly A^0 admits all paths, and by Lemma 4.3, if $s = 1, 2, \cdots, p + 1$, A^s admits all paths except $P(a_s, b_s)$.

THEOREM 4.1. *If A is defined by*

$$A_{ij} = \sum_{s=0}^{k} A^s(i, j) \qquad (0 \leqq k \leqq p + 1),$$

then A is the general square of prime order p admitting all paths $P(a_t, b_t)$ for which $k + 1 \leqq t \leqq p + 1$.

Proof. Obviously A admits the paths in question. Also A involves the variables $x_0, x_1, \cdots, x_{(p-1)k}$, and so has the right number of parameters, by Lemma 4.2. It merely remains to show that these parameters are independent. To do this it suffices to show that proper values can be assigned to the x's so as to make A any arbitrary square admitting the given paths. So let B admit the paths $P(a_t, b_t)$ for $k + 1 \leqq t \leqq p + 1$. Choose

$$x_0 = \frac{1}{p^2} \sum_{i,j} B_{ij}.$$

Then define C and D by

$$C_{ij} = B_{ij} - x_0, \qquad D_{ij} = A_{ij} - x_0.$$

That is,

$$D_{ij} = \sum_{s=1}^{k} A^s(i, j).$$

Clearly C admits the same paths that B does, and $N = 0$ for C. Also it is sufficient to show that $x_1, x_2, \cdots, x_{(p-1)k}$ may be chosen so that D and C are the same.

For $1 \leqq y \leqq p - 1$, $1 \leqq s \leqq k$, choose $x_{(p-1)(s-1)+y}$ as follows. Superpose

A^s on C; then those elements of A^s which consist of $x_{(p-1)(s-1)+\nu}$ alone cover a path $P(i_\nu, j_\nu; a_s, b_s)$ of C. Take

$$x_{(p-1)(s-1)+\nu} = \frac{1}{p} P(i_\nu, j_\nu; a_s, b_s).$$

With this choice for all the x's, it is clear that

$$-\sum_{\nu=1}^{p-1} x_{(p-1)(s-1)+\nu} = \frac{1}{p} P(0, 0; a_s, b_s).$$

That is, we can, and do, choose the x's in such a way that if $1 \leqq s \leqq k$, then each element of the path $P(i, j; a_s, b_s)$ in A^s equals

$$\frac{1}{p} \text{(the path } P(i, j; a_s, b_s) \text{ in } C).$$

It is then clear that

$$D_{ij} = \frac{1}{p} \sum_{s=1}^{k} P(i, j; a_s, b_s),$$

the $P(i, j; a_s, b_s)$ being all taken out of C. However, for C, $P(a_t, b_t) = 0$ if $k + 1 \leqq t \leqq p + 1$. So

$$D_{ij} = \frac{1}{p} \sum_{s=1}^{p+1} P(i, j; a_s, b_s).$$

That is, pD_{ij} is the sum of all the paths of C through C_{ij}. Each pair of these paths already has C_{ij} in common, and so by Lemma 4.3, no two of them have any other element in common. Hence the sum of all paths through C_{ij} contains C_{ij} $(p + 1)$ times and each other element of C_{ij} exactly once, and hence is

$$pC_{ij} + \sum_{i,j} C_{ij} = pC_{ij}.$$

This proves the theorem.

THEOREM 4.2. *A square of prime order admitting all paths but rows and columns is primitive.*[3]

Proof. Take $a_1 = 0, b_1 = 1, a_2 = 1, b_2 = 0, \alpha_1 = 1, \beta_1 = 0, \alpha_2 = 0, \beta_2 = 1$. Then A, defined by

$$A_{ij} = \sum_{s=0}^{2} A^s(i, j),$$

is clearly primitive.

Any transform of a p.s. shall be called a *regular square*.

COROLLARY 4.1. *A square of prime order admitting all but two paths is regular.*

Proof. By the right transformation, the two paths can be carried into rows

[3] This theorem is true for squares of any order. The proof is given in the supplement.

and columns, and the resulting square must be primitive by Theorem 4.2. Applying the inverse transformation proves the corollary.

COROLLARY 4.2. *Every d.s. of order 5 is regular.*

For a d.s. of order 5 admits all paths but two.

THEOREM 4.3. *The general d.s. of order 5 admits rows, columns, diagonals, the configuration $A(1, 1) + A(1, 2) + A(1, 5) + A(2, 1) + A(5, 1)$, and transforms of this latter under the group G of Theorem 3.3. It admits no other configuration which is the sum of five elements. The general d.s. S_p of prime order $\geqq 7$ admits no configuration C consisting of the sum of p elements, except rows, columns, and diagonals.*[4]

Proof. Take p a prime $\geqq 5$, let S_p be the general d.s. of order p, and let C be a configuration consisting of the sum of p elements. As the elements of A^s and A^t $(s \neq t)$ are independent, A^s must admit C if S_p does and $P(a_s, b_s)$ is not a row, column, or diagonal. Since C is the sum of p elements, A^s can admit C if and only if C contains exactly one element out of each path of the set $P(a_s, b_s)$. That is, the only paths of S_p which can contain two elements of C are rows, columns, or diagonals. Now let C not be a row, column, or diagonal and suppose S_p admits C. Take three elements of C which are not in the same row, column, or diagonal. We have just seen that each pair of elements out of C must lie in a row, column, or diagonal. Hence the three pairs of the three elements which we chose must lie respectively in either a row, column, and diagonal, or a row and one of each kind of diagonal, or a column and each kind of diagonal. In the two latter cases we can apply Q of Theorem 3.3 and say that the pairs lie in a row, column, and diagonal. By further transformations of G, we can take the pair which are in a row to be $A(1, 1)$ and $A(1, 2)$, with the third one lying in the column with $A(1, 1)$ and in a diagonal with $A(1, 2)$. Then the third is either $A(2, 1)$ or $A(p, 1)$. By a final transformation of G, we take the third one to be $A(2, 1)$. Now let $A(i, j)$ be any other element of C. Then $A(i, j)$ must be in a row, column or diagonal with each of $A(1, 1)$, $A(1, 2)$, and $A(2, 1)$. That is, $P(i - 1, j - 1)$, $P(i - 1, j - 2)$, and $P(i - 2, j - 1)$ must each be a row, column, or diagonal. By trying all possibilities, we see that either $i = 1, j = p$, or $i = p, j = 1$, or $i = j = 2$, or $i = j = \frac{1}{2}(p + 3)$. However, neither $A(2, 2)$ nor $A(\frac{1}{2}(p + 3), \frac{1}{2}(p + 3))$ is in the same row or column with either $A(1, p)$ or $A(p, 1)$. So C must consist of either $A(1, 1) + A(1, 2) + A(2, 1) + A(1, p) + A(p, 1)$, or $A(1, 1) + A(1, 2) + A(2, 1) + A(2, 2) + A(\frac{1}{2}(p + 3), \frac{1}{2}(p + 3))$. Applying $L^2M^{-1}Q$ of Theorem 3.3 to the latter, we get the former. Hence the latter is a transform of the

[4] The d.s. of order 5 admits the configuration named, the d.s. of order 4 admits $L(2)$ and $A_{00} + A_{01} + A_{10} + A_{11}$, the d.s. of order 8 admits $L(0, 0; 4) + L(2, 2; 4)$ and the d.s. of order 9 admits $L(3)$. So far as we know, these are the only values of n for which the general d.s. of order n admits a configuration of n elements other than a row, column, or diagonal.

former under the group G. This first-named configuration is admitted by S_5, as can be seen by inspection of the general square of order 5.

COROLLARY 4.3. *The only permutations of the elements of a general d.s. of prime order* $\geqq 7$ *which preserve its diabolicity are those which transform the set of rows, columns, and diagonals into rows, columns, and diagonals.*

COROLLARY 4.4. *The group of permutations of the general d.s. of prime order* $\geqq 7$ *into a d.s. is the group* G *of Theorem* 3.3, *of order* $8p^2(p-1)$.

THEOREM 4.4. *If* H *is the group generated by the transformation* P *of Theorem* 3.3 *and the permutations of the rows among themselves, and if*

$$T = \left\| \begin{matrix} 3 & 2 \\ 1 & 1 \end{matrix} \right\|,$$

then any transformation of the form TUT^{-1}, *where* U *is in* H, *will transform a d.s. of order* 5 *into a d.s. of order* 5.

Proof. If S_5 is a d.s., then $T^{-1}S_5$ is a p.s., $UT^{-1}S_5$ is a p.s., and $TUT^{-1}S_5$ is a d.s.

We will show in Corollary 5.2 that any transformation which carries the general d.s. of order 5 into a d.s. must have the form TUT^{-1}, where U is in H.

5. **Numerical squares.** Any square S_n whose elements are the integers $1, 2, \cdots, n^2$ is called a *numerical* square (n.s.). Thus we may speak of a numerical diabolic square (n.d.s.) or a numerical primitive square (n.p.s.). In a n.s. we have

$$N = \frac{1}{n^2}\sum_{i,j=1}^{n} A_{ij} = \frac{1}{n^2}\sum_{k=1}^{n^2} k = \frac{1}{n^2}\,\frac{n^2(n^2+1)}{2} = \frac{n^2+1}{2}.$$

THEOREM 5.1. *There are no n.d.s. of order* 3.

This follows immediately from Lemma 4.1.

THEOREM 5.2. *There are no n.d.s. of order* n *if* $n \equiv 2 \pmod 4$.

Proof. Let $n = 4m + 2$, and suppose S_n is a n.d.s. Then S_n admits the lattice $L(2)$ by Theorem 2.4. Since the elements of the square are integers, $L(2)$ is an integer. However, $L(2)$ has $(2m+1)^2$ elements, whose average value N is $\frac{1}{2}(n^2+1)$. So

$$L(2) = (2m+1)^2\tfrac{1}{2}(n^2+1),$$

which is not an integer.

In the course of showing that there are n.d.s. of all orders not excluded by the two preceding theorems, it is convenient to know the possible structure of n.p.s. This we now study.

Until Theorem 5.3, we assume that the arguments of $A(i, j)$ are always in the range $1, 2, \cdots, n$.

From the definition, the property of being primitive is invariant under any transformation of H in Theorem 4.4. Hence we may normalize our n.p.s. so that

$$(1)\qquad \begin{gathered} A(1, 1) = 1, \qquad A(1, 2) < A(2, 1), \qquad A(1, i) < A(1, i + j), \\ A(i, 1) < A(i + j, 1) \end{gathered}$$

if $1 \leqq j$. Let $\Phi(n)$ be the number of normalized n.p.s. of order n. Then the total number of n.p.s. is $2(n!)^2\Phi(n)$. From

$$(2)\qquad A_{ij} + A_{kl} = A_{il} + A_{kj}$$

and (1), it follows that

$$(3)\qquad A(i, j) < A(i + k, j + l)$$

if $0 \leqq k$, $0 \leqq l$, $1 \leqq k + l$. From (3), we see that $A(1, 2) = 2$. Let m be the greatest value of j such that for $i \leqq j$, $A(1, i) = i$. Then $2 \leqq m \leqq n$; also, from (3), we have $A(2, 1) = m + 1$. We wish to show first that the elements occur in the rows in sets of m consecutive integers. If this is true of the first row, it follows for the others by (2). So suppose it is false for the first row. Let the first rm elements of the first row occur in such sets, but let

$$A(1, mr + 1), A(1, mr + 2), \cdots, A(1, m(r + 1))$$

not be consecutive integers. Let k be the largest integer such that for all j with $1 \leqq j \leqq k$

$$A(1, mr + j) = A(1, mr + 1) - 1 + j.$$

Then $1 \leqq k \leqq m - 1$. Then $A(1, mr + 1) + k$ does not occur in the first row. Let it occur in the i-th row.

Case 1. There is an $l \geqq 1$ such that

$$A(1, mr + 1) + k = A(i, mr + l).$$

Since $i \geqq 2$, we have

$$\begin{aligned} A(1, mr + 1) + k &= A(i, mr + l) \geqq A(i, mr + 1) \\ &= A(1, mr + 1) + A(i, 1) - A(1, 1) \geqq A(1, mr + 1) + A(2, 1) - A(1, 1) \\ &\geqq A(1, mr + 1) + m + 1 - 1 \geqq A(1, mr + 1) + m. \end{aligned}$$

This contradicts the condition $k \leqq m - 1$.

Case 2. There is a j $(0 \leqq j \leqq r - 1)$ and an l $(1 \leqq l \leqq m)$ such that

$$A(1, mr + 1) + k = A(i, mj + l)$$

Then since the elements of the first rm columns occur in the rows in sets of m consecutive integers,

$$A(1, mr + 1) + k + w - l = A(i, mj + w)$$

for $1 \leqq w \leqq m$. If $l \geqq 2$, then $A(1, mr + 1) + k - 1$ must occur in both the first and i-th rows. So $l = 1$. Hence

$$A(1, mr + 1) + m = A(i, mj + m + 1 - k).$$

However,

$$A(1, 1) + A(2, mr + 1) = A(1, mr + 1) + A(2, 1),$$

and $A(1, 1) = 1$, $A(2, 1) = m + 1$, so that

$$A(2, mr + 1) = A(1, mr + 1) + m = A(i, mj + m + 1 - k),$$

and the conditions $k \geqq 1$, $j \leqq r - 1$ are contradicted.

So we conclude that the elements occur in the rows in sets of m consecutive integers. Then clearly the elements $A(i, mj)$ must be multiples of m. If we define B by

$$B(i, j) = \frac{1}{m} A(i, mj),$$

then clearly the rectangle B is primitive, and

$$B(1, 1) = 1, \qquad B(2, 1) = 2, \qquad B(i, j) \leqq B(i + k, j + l)$$

if $0 \leqq k$, $0 \leqq l$, $1 \leqq k + l$. Since the preceding arguments would hold equally well for rectangles, we may interchange the rows and columns of B and repeat the argument. This process evidently terminates after a finite number of steps. Thus the normalized n.p.s. of order four may have the structures

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

1	2	9	10
3	4	11	12
5	6	13	14
7	8	15	16

1	2	5	6
3	4	7	8
9	10	13	14
11	12	15	16

and no others.

An important property of a n.p.s. is the fact that the elements 1 and 2 always occur in the same row or column. This property shall be of use in the construction of irregular squares. The exact number $\Phi(n)$ of normalized n.p.s. of order n is the number of ways of choosing $d_1, d_2, \cdots, d_s$ and $f_1, f_2, \cdots, f_t$ so

that $s = t$ or $t + 1$; $d_1 \neq 1$, $d_s = n$; $f_1 \neq 1$, $f_t = n$; $d_i \neq d_{i+1}$, $f_i \neq f_{i+1}$; d_i divides d_{i+1}; and f_i divides f_{i+1}. This is certainly less than or equal to $(n!)^2$, since d_1 cannot be chosen in more than n ways, d_2 cannot be chosen in more than $n - 1$ ways, etc., and the same applies to the f's.

For prime orders there is clearly exactly one normalized n.p.s. Hence there are exactly two types of n.p.s. of prime order (cf. the discussion following the proof of Theorem 3.1), namely, those of the same type as A, where

$$A_{ij} = (i - 1)p + j,$$

and those of the same type as PA (P as in Theorem 3.3). If B is got from PA by a regular transformation T, then B is got from A by the regular transformation TP. A n.d.s. cannot be got from either A or PA by an irregular transformation. Such a transformation of any n.p.s. of prime order must carry either rows or columns into rows, columns, or diagonals, and since no two rows or columns of a n.p.s. have the same sum, the transformed square would not be diabolic. Hence a n.d.s. of prime order is regular if and only if it can be got from a n.p.s. of the type of A by a regular transformation, and so by Theorem 3.2, we have

THEOREM 5.3. *There are exactly* $(p!)^2(p - 3)(p - 4)$ *regular n.d.s. of order* p *if* p *is an odd prime.*

The regular n.d.s. of odd prime order obtained from normalized n.p.s. are the so-called "step squares". The regular n.d.s. are the natural generalization of "step squares". For instance, the generalized "step squares" discussed by A. H. Frost[6] are all regular n.d.s.

COROLLARY 5.1. *There are precisely* 28,800 *n.d.s. of order* 5.

Use the above theorem and Corollary 4.2.

COROLLARY 5.2. *Any transformation which carries the general diabolic square of order* 5 *into a diabolic square must have the form* TUT^{-1}, *where* T *and* U *are as in Theorem* 4.4.

Proof. Any transformation V which carries the general d.s. of order 5 into a d.s. must carry a n.d.s. into a n.d.s. Hence there are at least as many n.d.s. as there are transformations V. So there are at most 28,800 transformations V. However, TUT^{-1} is clearly a V. Also H is of order 28,800 $(= 2(5!)^2)$ since one can arrange the rows in 5! ways, the columns in 5! ways, and orient the square in two ways. So there are exactly 28,800 V's and each one is a TUT^{-1}.

Since n.p.s. of all orders exist, the existence of regular n.d.s. of any order prime to 6 follows from Theorem 3.2. We shall construct regular n.d.s. of all orders except those excluded by Theorems 5.1 and 5.2. We shall also construct irregular n.d.s. of all orders except those excluded by Theorems 5.1, 5.2, and 5.4.

[6] A. H. Frost, *On the general properties of Nasik squares*, Quart. Jour. Math., vol. 15 (1877), pp. 34–49.

THEOREM 5.4. *Any d.s. of order 1, 4 or 5 must be regular.*

Proof. The theorem is obvious for $n = 1$. For $n = 5$, use Corollary 4.2. For a square of order 4, one can find the general square simply by solving the equations which say that the square admits rows, columns, and diagonals. The general square is then seen to be the transform by

$$\begin{Vmatrix} 1 & 2 \\ 2 & 1 \end{Vmatrix}$$

of a p.s.

LEMMA 5.1. *If m and n are both odd integers greater than or equal to three, it is possible to arrange the integers $0, 1, 2, \cdots, mn - 1$ in a rectangular array of n rows and m columns in such a way that the sum of the n elements in each column is the same.*

Proof. Let the first row start with 0 in the first column and increase by a unit at a time to $m - 1$ in the m-th column. Let the second row start with $\frac{1}{2}(3m + 1)$ in the first column and increase by a unit at a time to $2m - 1$ in the $(\frac{1}{2}(m - 1))$-th column, and then start with m in the $(\frac{1}{2}(m + 1))$-th column and increase by a unit at a time to $\frac{1}{2}(3m - 1)$ in the m-th column. Let the third row start with $3m - 2$ in the first column and decrease by two units at a time to $2m + 1$ in the $(\frac{1}{2}(m - 1))$-th column and then start with $3m - 1$ in the $(\frac{1}{2}(m + 1))$-th column and decrease by two units at a time to $2m$ in the m-th column. Clearly the sum of the first three elements in each column is $\frac{1}{2}(9m - 3)$. If $n \geqq 5$, the array may be completed by starting even-numbered rows with multiples of m in the first column and increasing a unit at a time to the right, and by starting odd-numbered rows with multiples of m in the m-th column and increasing a unit at a time to the left.

THEOREM 5.5. *There are regular n.d.s. of all orders except those excluded by Theorems* 5.1 *and* 5.2.

Proof. Case 1. If n is prime to 6, use Theorem 3.2.

Case 2. Let $n = 4m$, and let

$$T = \begin{Vmatrix} 2 & -1 \\ -3 & 2 \end{Vmatrix}.$$

Put $a(i) = i$, $a(i + 2m) = 4m + 1 - i$ for $1 \leqq i \leqq 2m$, and put $a(i \pm 4m) = a(i)$, and define A by

$$A_{ij} = na(i) + a(j) - n.$$

Then A is clearly a n.p.s. We will show that TA is a n.d.s., which is therefore regular. T transforms the paths $P(1, 2)$, $P(2, 3)$, $P(3, 5)$, $P(1, 1)$ of A into the paths $P(0, 1)$, $P(1, 0)$, $P(1, 1)$, and $P(1, -1)$. So we merely need to show that A admits the first set of paths named. The proofs are similar for each path, and we will illustrate them by giving the proof for $P(2, 3)$.

$$P(i, j; 2, 3) = \sum_{x=1}^{n} (na(i + 2x) + a(j + 3x) - n)$$

$$= n \sum_{x=1}^{n} a(i + 2x) + \sum_{x=1}^{n} a(j + 3x) - n^2.$$

Now

$$\sum_{x=1}^{n} a(i + 2x) = 2 \sum_{x=1}^{2m} a(i + 2x) = 2 \sum_{x=1}^{m} [a(i + 2x) + a(i + 2x + 2m)]$$

$$= 2 \sum_{x=1}^{m} (4m + 1) = 2m(4m + 1) = \tfrac{1}{2}n(n + 1).$$

If $m \not\equiv 0 \pmod 3$,

$$\sum_{x=1}^{n} a(j + 3x) = \sum_{x=1}^{n} a(x) = \tfrac{1}{2}n(n + 1).$$

If $m = 3s$,

$$\sum_{x=1}^{n} a(j + 3x) = 3 \sum_{x=1}^{4s} a(j + 3x) = 3 \sum_{x=1}^{2s} [a(j + 3x) + a(j + 3x + 2m)]$$

$$= 3 \cdot 2s(n + 1) = \tfrac{1}{2}n(n + 1).$$

In either case,

$$P(i, j; 2, 3) = n\tfrac{1}{2}(n^2 + 1) = nN.$$

Case 3. $n = 3m$, m odd, $m \geqq 3$. Let

$$T = \begin{Vmatrix} 1 & 1 \\ 2 & 3 \end{Vmatrix}.$$

Arrange the numbers $0, 1, 2, \cdots, 3m - 1$ in the array of Lemma 5.1, taking 3 columns and m rows. Put $a(3x + y)$ equal to the element in the $(x + 1)$-th row and y-th column of the array and put $a(i \pm n) = a(i)$. Then

$$\sum_{x=1}^{m} a(3x + y) = \tfrac{1}{6}n(n - 1).$$

Define A by

$$A_{ij} = na(i) + a(j) + 1.$$

A is a n.p.s. T transforms the paths $P(1, -1)$, $P(3, -2)$, $P(2, -1)$, $P(4, -3)$ into $P(0, 1)$, $P(1, 0)$, $P(1, 1)$, $P(1, -1)$. A admits $P(1, -1)$ and $P(2, -1)$ by Theorem 2.1. We prove that A admits $P(4, -3)$, and the proof that A admits $P(3, -2)$ is similar.

$$P(i, j; 4, -3) = \sum_{x=1}^{n} (na(i + 4x) + a(j - 3x) + 1)$$

$$= n \sum_{x=1}^{n} a(i + 4x) + \sum_{x=1}^{n} a(j - 3x) + n.$$

Now

$$\sum_{x=1}^{n} a(i + 4x) = \sum_{x=1}^{n} a(x) = \tfrac{1}{2}n(n - 1),$$

and

$$\sum_{x=1}^{n} a(j - 3x) = 3 \sum_{x=1}^{m} a(j + 3x) = \tfrac{1}{2}n(n - 1).$$

Hence $P(i, j; 4, -3) = nN$.

THEOREM 5.6. *There are non-regular n.d.s. of all orders except those excluded by Theorems* 5.1, 5.2, *and* 5.4.

Case 1. n odd and $(n, 6) = 1$.

2	47	38	35	24	20	9
26	16	8	6	46	42	31
49	39	33	23	15	12	4
19	11	7	45	41	30	22
37	29	27	17	14	3	48
10	5	44	36	34	25	21
32	28	18	13	1	43	40

is a non-regular n.d.s. One can see from the discussion preceding Theorem 5.3 that if it were regular it would have to be the transform of a n.p.s. in which one column contained 1, 8, 15, 22, 29, 36, 43. But these do not constitute a path in the given square, and hence such a transformation is impossible.

We now take $n \geqq 11$. Let A be defined by

$$A_{ij} = (i - 1)n + a_j,$$

where $a_1, a_2, \cdots, a_{11} = 2, 7, 4, 9, 6, 11, 1, 8, 3, 10, 5$ respectively, and $a_j = j$ for $12 \leqq j \leqq n$. Define B as the square with $B(1, 4)$, $B(4, 4)$, $B(4, 2)$, $B(3, 4)$, $B(3, 9)$, $B(2, 11)$, $B(2, 9)$, $B(5, 9)$, $B(\frac{1}{2}(n + 7), 2)$, $B(\frac{1}{2}(n + 5), 4)$, $B(\frac{1}{2}(n + 5), 2)$, $B(\frac{1}{2}(n + 11), 2)$, $B(\frac{1}{2}(n + 3), 9)$, $B(\frac{1}{2}(n + 9), 9)$, $B(\frac{1}{2}(n + 9), 7)$, and $B(\frac{1}{2}(n + 7), 9)$ equal to $+1$; $B(3, 3)$, $B(2, 5)$, $B(2, 3)$, $B(5, 3)$, $B(1, 10)$, $B(4, 10)$, $B(4, 8)$, $B(3, 10)$, $B(\frac{1}{2}(n + 3), 3)$, $B(\frac{1}{2}(n + 9), 3)$, $B(\frac{1}{2}(n + 9), 1)$, $B(\frac{1}{2}(n + 7), 3)$, $B(\frac{1}{2}(n + 7), 8)$, $B(\frac{1}{2}(n + 5), 10)$, $B(\frac{1}{2}(n + 5), 8)$, and $B(\frac{1}{2}(n + 11), 8)$ equal to -1; and all other elements equal to zero. Put

$$T = \begin{Vmatrix} 2 & 1 \\ 1 & 2 \end{Vmatrix}.$$

Then TA is a n.d.s., since T is regular and A is a n.p.s. Also it can be readily seen that each row, column, or diagonal of TB contains as many $+1$'s as -1's, so that TB is diabolic. Hence $TA + TB$ is diabolic. Hence it is a n.d.s. if it contains each of the elements $1, 2, \cdots, n^2$. However, $TA + TB = T(A + B)$, and $(A + B)$ is readily seen to be the result of interchanging pairs of elements $A_{i,j}$ and $A_{i,j+6}$ in A. So $TA + TB$ is a n.d.s. If $T(A + B)$ were regular, $A + B$ would also be regular. We have $A_{17} + B_{17} = 1$, $A_{11} + B_{11} = 2$; these elements lie on the unique path $P(1, 1; 0, 6) = P(1, 1; 0, 1)$ since $7 - 1 = 6$ is prime to n, and so any transform of $A + B$ into a p.s. would take the entire first row of $A + B$ into a row. Since this row consists of the numbers $1, \cdots, n$, the number $n + 1$ must lie in the same column of the transformed square as 1. Since $n + 1 = A_{27} + B_{27}$ lies in the same column of $A + B$ as 1, the transformation would take rows into rows and columns into columns, so that if the transformed square were primitive, $A + B$ would be primitive. But this is obviously not so.

Case 2. $n = 4m, m \geqq 2$. Consider the $2m^2$ pairs $(0, 16m^2 - 4), (4, 16m^2 - 8), (8, 16m^2 - 12), \cdots, (8m^2 - 4, 8m^2)$. If (A, C) and (B, D) are two of the pairs, then

(4)

$A + 1$	$B + 4$	$C + 1$	$D + 4$
$C + 2$	$D + 3$	$A + 2$	$B + 3$
$A + 4$	$B + 1$	$C + 4$	$D + 1$
$C + 3$	$D + 2$	$A + 3$	$B + 2$

is diabolic with $N = \frac{1}{2}(16m^2 + 1)$. Fill in the $L(m)$'s of S_{4m} with squares like (4), using different sets of pairs for each square. Then the result is a n.d.s. by Theorem 2.7. The number of ways of assigning a pair of pairs to each $L(m)$ is

$$\frac{(2m^2)!}{2^{m^2}}.$$

The square composing any $L(m)$ can be arranged in 384 ways.[6] So there are at least

$$(192)^{m^2}(2m^2)! \tag{5}$$

n.d.s. of order $4m$.[7] We will now show that this exceeds the number of regular n.d.s. of order $4m$. Let $\psi(4m)$ be the number of normalized n.p.s. of order $4m$. Then there are $2((4m)!)^2\psi(4m)$ n.p.s. of order $4m$. However, by an argument

[6] Barkley Rosser and R. J. Walker, loc. cit., Theorem 4.

[7] The squares of order $4m$ constructed by this method satisfy the conditions that each $L(m)$ is a d.s. of order 4. This gives $12m^2$ independent linear conditions on the elements. As the general square of order $4m$ satisfies only $16m - 4$ independent linear conditions, the number of such squares must be immense.

like that preceding Theorem 5.3, we see that all regular n.d.s. are transforms of only $((4m)!)^2\psi(4m)$ n.p.s. Let

$$T = \begin{Vmatrix} a & c \\ b & d \end{Vmatrix}.$$

By a discussion like that preceding Lemma 3.1, we see that if either b or d is prime to $4m$, it can be taken as unity. So for $m = 2$, we may consider the following cases: $b = d = 1$; $b = 1$, d even; b even, $d = 1$; b and d both even. For each case, a and c can be chosen in at most 8 ways each. So there are at most $64 + 4\cdot 64 + 4\cdot 64 + 16\cdot 64 = 1600$ transformations which could yield different n.d.s. of order 8. As $\psi(8) = 10$, we see that the number of regular n.d.s. of order 8 is less than the number of n.d.s. of order 8 given by (5). For $m \geqq 3$, we may simply take the number of transformations as less than $(4m)^4$. Hence there are at most

$$(4m)^4((4m)!)^2\psi(4m) \tag{6}$$

regular n.d.s. of order $4m$. $\psi(12) = 42$, so that for $m = 3$, (6) is less than (5). We have shown that $\psi(n) \leqq (n!)^2$. Using this in (6) and using approximate values of $n!$, we see that for $m \geqq 4$, (5) exceeds (6).

Case 3. $n = 9$.

(7)

1	65	48	41	24	58	81	34	17
57	76	33	13	8	70	53	39	20
68	49	44	27	62	75	28	12	4
80	36	16	3	64	47	40	23	60
52	38	19	56	78	32	15	7	72
30	11	6	67	51	43	26	61	74
42	22	59	79	35	18	2	66	46
14	9	71	54	37	21	55	77	31
25	63	73	29	10	5	69	50	45

is a non-regular n.d.s. For if it were regular, one of the paths containing 1 and 2 would be the transform of a row of a n.p.s. But the only paths containing 1 and 2 are $P(1, 1)$, $P(1, 4)$, and $P(1, 7)$. Since (7) admits $P(1, 1)$, these paths cannot be the rows of a n.p.s. The elements of $P(1, 1; 1, 4)$ are 1, 2, 3, 4, 5, 6, 7, 8, 9. If this were a row of a n.p.s., the numbers 1, 10, 19, 28, 37, 46, 55, 64, 73 would have to form a column. However, these do not constitute a path in (7). The elements of $P(1, 1; 1, 7)$ are 1, 2, 3, 37, 38, 39, 73, 74, 75. These cannot constitute a row of a n.p.s.

Case 4. $n = 3m$, $(m, 6) = 1$, $m \geqq 5$. If $a(x, 1), a(x, 2), \cdots, a(x, m)$ are a set of m distinct integers lying between 0 and $9m - 1$ inclusive, and if A_x is defined by

$$A_{ij} = 9mi + a(x, j) - 9m + 1,$$

then by Theorem 3.2 there is a regular transformation T such that TA_x is a d.s. Arrange the numbers $0, 1, 2, \cdots, 9m - 1$ in the array of Lemma 5.1, taking m rows and 9 columns. Take $a(x, y)$ to be the element in the y-th row and x-th column. Then $TA_1, TA_2, \cdots, TA_9$ are each d.s., and together use the integers $1, 2, \cdots, n^2$. So if we take the TA_x's as lattices $L(3)$ of a square of order n, it will be a n.d.s. by Theorem 2.7. The lattices may be inserted in such a way that 1, 2, and 3 occur as the first three elements of the first row of the n.d.s. of order n. Then by inspection of the array of Lemma 5.1, it is seen that the remaining elements of the first row are $9m\alpha + 15$, $9m\alpha + 16$, $9m\alpha + 17$, $9m\beta + 22$, $9m\beta + 24$, $9m\beta + 26$, $9m\gamma + 28$, $9m\gamma + 29$, $9m\gamma + 30, \cdots$. Any transformation taking the square into a p.s. would have to take the first row into a row, since it is the only path containing 1 and 2. But the row consists of $m - 1$ triplets of consecutive integers and one triplet of non-consecutive integers and this cannot be a row of a n.p.s. Hence the square in question is irregular.

Case 5. $n = 9m$, m odd, $m \geqq 3$. Take T as in Case 2 of Theorem 5.5. Arrange the numbers $0, 1, 2, \cdots, 27m - 1$ in the array of Lemma 5.1, taking 27 columns and m rows. Put $a(x, 3y + z)$ equal to the element in the $(y + 1)$-th row and the $(x + 9z - 9)$-th column $(x = 1, 2, \cdots, 9)$. Define A_x by

$$A_{ij} = 27mi + a(x, j) - 27m + 1.$$

By an argument like that of Case 2, Theorem 5.5, TA_x is a d.s. Then we can use the TA_x's as lattices $L(3)$ of a n.d.s. of order n. The argument of the preceding case will show that this can be done in such a way that the n.d.s. of order n is irregular. This completes the proof of the theorem.

If A and B are n.d.s. of orders m and n, and we define the lattice $L(i, j; m)$ of a square D of order mn to be

$$C_{xy} = B_{xy} + n(A_{ij} - 1),$$

then D is a n.d.s. by Theorem 2.7. If B is irregular, then D is irregular by Theorem 2.2 and the fact that lattices are transformed into lattices. If B is regular, one can transform some but not all of the lattices of D by transformations of the group G of Theorem 3.3, and the result will usually be irregular.

In the supplement is described a means of constructing a n.d.s. B of order $3m$ from a n.d.s. A of order m. It is proved that if A is irregular, then B is irregular.

6. **Conclusion.** The methods used in obtaining properties of diabolic squares can be applied to other similar problems. It is evident that diabolic cubes (or hypercubes) can be defined and studied in an analogous manner. In the supplement is established the existence or nonexistence of a numerical diabolic cube of any given order.

Another application is to the theory of latin squares. A latin square is most easily defined as a square whose elements are n independent variables each occurring n times. Thus a diabolic latin square (d.l.s.) must contain each variable once in each row, column, and diagonal. The principal results concerning d.l.s. proved in the supplement are the following.

(a) There are no d.l.s. of order divisible by 2 but not by 8, or by 3 but not by 9.

(b) There are regular d.l.s. of order n if and only if $(n, 6) = 1$.

(c) There are irregular d.l.s. of prime order p if and only if $p > 11$.

CORNELL UNIVERSITY.

A MATRIX THEORY OF n-DIMENSIONAL MEASUREMENT

BY EVERETT H. LARGUIER, S.J.

1. **Introduction.** In 1933 A. H. Copeland[1] gave a precise statement of the fundamental assumptions of a theory of measurement. It would serve for a one-dimensional theory of probability measurement. It is our purpose here to extend some of his results to n dimensions, thereby simplifying the treatment of n-dimensional probability theory.

With Copeland, a matrix of the form $\mathbf{x} = x^{(1)}, x^{(2)}, \cdots, x^{(k)}, \cdots$, where the k-th term $x^{(k)}$ is an arbitrary number, will be called a *variate*. If $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ is a set of n variates, and if $G(s_1, s_2, \cdots, s_n)$ is an arbitrary function of the variables $s_1, s_2, \cdots, s_n$, then $G(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ denotes a variate defined as follows: $G(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n) = G(x_1^{(1)}, x_2^{(1)}, \cdots, x_n^{(1)}), G(x_1^{(2)}, x_2^{(2)}, \cdots, x_n^{(2)}), \cdots$. We shall define the probable value of the matrix $\mathbf{x}$ as

$$\mathbf{p}(\mathbf{x}) = \lim_{n\to\infty} \mathbf{p}_n(\mathbf{x}), \quad \text{where} \quad \mathbf{p}_n(\mathbf{x}) = \sum_{k=1}^{n} \frac{x^{(k)}}{n}.$$

Sometimes, however, this value $\mathbf{p}(\mathbf{x})$ may not exist, as is obvious.

In a similar manner we shall understand by an *n-dimensional variate* a variate defined by the matrix

$$(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n) = (x_1^{(1)}, x_2^{(1)}, \cdots, x_n^{(1)}), (x_1^{(2)}, x_2^{(2)}, \cdots, x_n^{(2)}), \cdots$$

in which the k-th term $(x_1^{(k)}, x_2^{(k)}, \cdots, x_n^{(k)})$ designates a point in an n-dimensional space. Further we will understand by $\phi_E(s_1, s_2, \cdots, s_n)$ the fundamental function of the point set E, and this function will have the value 1 or 0 according as the point $(s_1, s_2, \cdots, s_n)$ is a point of E or not. Then

$$\mathbf{p}[\phi_E(\mathbf{x}_1, \cdots, \mathbf{x}_n)] = \lim_{n\to\infty} \mathbf{p}_n[\phi_E(\mathbf{x}_1, \cdots, \mathbf{x}_n)],$$

where $\mathbf{p}_n[\phi_E(\mathbf{x}_1, \cdots, \mathbf{x}_n)]$ is defined as above. Hence if we let

$$\mathbf{p}[\phi_E(\mathbf{x}_1, \cdots, \mathbf{x}_n) = F(S_1, S_2, \cdots, S_n)$$

in which E is the n-dimensional cell defined as follows: $-\infty \leqq s_i \leqq S_i$ ($i = 1, 2, \cdots, n$), then $F(S_1, S_2, \cdots, S_n)$ will be the probability that the so-called digits of the matrix $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ will designate a point in the cell E. This function $F(s_1, s_2, \cdots, s_n)$ will be called the *n-dimensional accumulative proba-*

Received November 12, 1938; presented to the American Mathematical Society, December 30, 1938. The author wishes to acknowledge his indebtedness to Professor Francis Regan, of Saint Louis University, for the assistance and guidance he gave during the preparation of this paper.

[1] *A matrix theory of measurement*, Math. Zeit., vol. 37 (1933), pp. 542–555.

bility function. It will be noted that $F(-\infty, -\infty, \cdots, -\infty) = 0$ and $F(+\infty, +\infty, \cdots, +\infty) = 1$. The function is monotonic non-decreasing with respect to each variable, that is to say, $F(s_1, s_2, \cdots, s_n) \geqq F(s_1', s_2', \cdots, s_n')$, where $s_k > s_k'$ and $s_i = s_i'$, $i \neq k$ for all subscripts $1, 2, \cdots, n$. Moreover, when $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ is a set of independent variates,[2] then

$$F(s_1, s_2, \cdots, s_n) = F_1(s_1)F_2(s_2)F_3(s_3) \cdots F_n(s_n),$$

where $F_i(s_i) = \mathbf{p}[\phi_{E_i}(\mathbf{x}_i)]$, E_i being the interval $(-\infty, s_i)$ in the n-dimensional space. Such an n-dimensional variate will be called an *independent n-dimensional variate*. If a set of n-dimensional variates $(\mathbf{x}_{11}, \mathbf{x}_{12}, \cdots, \mathbf{x}_{1n})$, $(\mathbf{x}_{21}, \mathbf{x}_{22}, \cdots, \mathbf{x}_{2n})$, $\cdots$, $(\mathbf{x}_{k1}, \mathbf{x}_{k2}, \cdots, \mathbf{x}_{kn})$ is such that

$$\mathbf{p}[\phi_E\{(\mathbf{x}_{11}, \cdots, \mathbf{x}_{1n})(\mathbf{x}_{21}, \cdots, \mathbf{x}_{2n}) \cdots (\mathbf{x}_{k1}, \cdots, \mathbf{x}_{kn})\}] = \mathbf{p}[\phi_E(\mathbf{x}_{11}, \cdots, \mathbf{x}_{1n})] \cdots \mathbf{p}[\phi_E(\mathbf{x}_{k1}, \cdots, \mathbf{x}_{kn})],$$

then the set will be termed an *independent set of n-dimensional variates*. This should not be confused with a *set of independent n-dimensional variates* which does not necessarily have this property.

The condition for n-dimensional admissibility may be expressed as follows:

An n-dimensional variate $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ *is said to be admissible with respect to a given function, provided*[3]

$$\mathbf{p}\left[\prod_{i=1}^{k}\left(\frac{r_i}{m}\right)\phi_{J_i}(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)\right] = \prod_{i=1}^{k}\int_{J_i} d_nF(s_1 + 0, s_2 + 0, \cdots, s_n + 0)$$

for every set of integers $r_1, r_2, \cdots, r_k, m$ *such that* $0 < r_1 < r_2 < \cdots < r_k \leqq m$ *and for every set of finite n-dimensional cells* J_i *each of which has the property*

$$J_i\colon\ a_{ij} < s_j \leqq b_{ij} \qquad (i = 1, 2, \cdots, k; j = 1, 2, \cdots, n)$$

In the case that $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ is an independent n-dimensional variate, this condition becomes

$$\mathbf{p}\left[\sum_{i=1}^{k}\left(\frac{r_i}{m}\right)\phi_{J_i}(\mathbf{x}_1, \cdots, \mathbf{x}_n)\right] = \prod_{i=1}^{k}\int_{I_{1i}}\int_{I_{2i}}\cdots\int_{I_{ni}} dF_1(s_1 + 0) \cdots dF_n(s_n + 0),$$

where r_i and m are as above and $F_j(s_j)$ is the accumulative distribution function of the variate $\mathbf{x}_j$ and I_{ij} is any set of finite intervals on the s_j-axis, open on the left and closed on the right, i.e., $I_{ij}\colon\ a_{ij} < s_j \leqq b_{ij}$.

The variation $\Delta_{a_1, a_2, \cdots, a_n}^{b_1, b_2, \cdots, b_n} F(s_1, s_2, \cdots, s_n)$ shall be defined by the following expression:[4]

[2] Copeland, ibid., p. 543.

[3] For the meaning of this integral the reader may consult M. Fréchet, *Extension au cas des intégrales multiples d'une définition de l'intégrale due à Stieltjes*, Nouv. Ann. de Math., (4), vol. 10(1910), pp. 241–256. The operator (r/m) is such that if $\mathbf{x} = x^{(1)}, x^{(2)}, \cdots$, then $(r/m)\mathbf{x} = x^{(r)}, x^{(r+m)}, x^{(r+2m)}, \cdots$.

[4] Fréchet, op. cit., p. 242; also E. W. Hobson, *Theory of Functions of a Real Variable*, vol. 1, pp. 343 ff.

$$
\begin{aligned}
&F(b_1, b_2, \cdots, b_n) \\
&\quad - [F(a_1, b_2, \cdots, b_n) + F(b_1, a_2, b_3, \cdots, b_n) + \cdots + F(b_1, \cdots, b_{n-1}, a_n)] \\
&\quad + [F(a_1, a_2, b_3, \cdots, b_n) + \cdots + F(b_1, b_2, \cdots, b_{n-2}, a_{n-1}, a_n)] \\
&\quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\
&\quad + (-1)^{n-1}[F(b_1, a_2, \cdots, a_n) + \cdots + F(a_1, a_2, \cdots, a_{n-1}, b_n)] \\
&\quad + (-1)^n F(a_1, a_2, \cdots, a_n).
\end{aligned}
$$

2. **Existence and properties of n-dimensional variates.** By a theorem of Copeland[5] there exists a set of points D which is admissibly ordered with respect to the function $\Pi(E)$ (the *Lebesgue n-dimensional measure* of E) defined in the unit cell

$$\Delta\colon\ 0 \leqq y_i < 1 \qquad (i = 1, 2, \cdots, n).$$

A simple method for establishing the existence of an n-dimensional variate with respect to a given distribution function is to set up a transformation in terms of the distribution function carrying the admissibly ordered points of the unit-cell into points of the whole space. To simplify the discussion we will first treat the case of a two-dimensional variate.[6]

It is desired to set up a transformation which carries the points (y_1, y_2) of the unit square into the points (s_1, s_2) of the whole plane. We shall first divide the unit square into strips by drawing vertical lines. These strips R_n will consist of points satisfying the inequalities

$$F(\infty, n-1) \leqq y_1 < F(\infty, n) \qquad (0 \leqq y_2 < 1). \tag{2.1}$$

It will be noted that each strip of the unit square corresponds to a strip of unit width and infinite length in the plane. We next divide the strips R_n into rectangles R_{mn} by drawing horizontal lines such that the points of the rectangles R_{mn} will satisfy the inequalities

$$
\begin{gathered}
F(\infty, n-1) \leqq y_1 < F(\infty, n), \\
\frac{F(m-1, n) - F(m-1, n-1)}{F(\infty, n) - F(\infty, n-1)} \leqq y_2 < \frac{F(m, n) - F(m, n-1)}{F(\infty, n) - F(\infty, n-1)}.
\end{gathered} \tag{2.2}
$$

In this way the points of each rectangle R_{mn} in the unit square correspond to the points of a rectangle in the whole plane. Next we subdivide the rectangles of the unit square by constructing further vertical and horizontal lines determined by a new set of inequalities similar to (2.2). It is to be observed that, at each stage of such divisions, there is a correspondence between the rectangles of the unit square and the rectangles of the whole plane. This process of division

[5] Copeland, *Admissible numbers in the theory of geometrical probability*, Amer. Journ. of Math., vol. 53(1931), pp. 153–162, Theorem 5.

[6] This transformation was communicated to the author by Professor A. H. Copeland, who kindly examined the first draft of this paper. It obviates a serious difficulty in the author's original transformation.

is continued indefinitely. And in the limit R_{mn} of the unit square defines the point (y_1, y_2) and the corresponding rectangle of the whole plane defines (s_1, s_2) in the limit. Hence in this manner we set up a one-to-one correspondence between the points of the unit square and the points of the whole plane; that is to say, we have defined a transformation carrying the points of the unit square into the points of the plane.

Hence, corresponding to the set of points of the unit square which is admissibly ordered with respect to $\Pi(E)$, we can have by means of the transformation a 2-dimensional variate

$$(\mathbf{x}_1, \mathbf{x}_2) = (x_1^{(1)}, x_2^{(1)}), (x_1^{(2)}, x_2^{(2)}), \cdots$$

which is admissibly ordered with respect to the function $F(s_1, s_2)$.

In a similar manner we are able to define a transformation for the case of the n-dimensional variate. For the first division we satisfy the following inequalities, similar to (2.2):

$$F(\infty, \infty, \cdots, \infty, k_n - 1) \leqq y_1 < F(\infty, \infty, \cdots, \infty, k_n),$$

$$\frac{F(\infty, \cdots, \infty, k_{n-1} - 1, k_n) - F(\infty, \cdots, \infty, k_{n-1} - 1, k_n - 1)}{F(\infty, \cdots, \infty, k_n) - F(\infty, \cdots, \infty, k_n - 1)}$$
$$\leqq y_2 < \frac{F(\infty, \cdots, \infty, k_{n-1}, k_n) - F(\infty, \cdots, \infty, k_n - 1)}{F(\infty, \cdots, \infty, k_n) - F(\infty, \cdots, \infty, k_n - 1)},$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$

$$\frac{F(\infty, \cdots, \infty, k_{n-i+1} - 1, k_{n-i+2}, \cdots, k_n) - F(\infty, \cdots, \infty, k_{n-i+1} - 1, \cdots, k_n - 1)}{F(\infty, \cdots, \infty, k_{n-i+2}, \cdots, k_n) - F(\infty, \cdots, \infty, k_{n-i+2} - 1, \cdots, k_n - 1)}$$
$$\leqq y_i < \frac{F(\infty, \cdots, \infty, k_{n-i+1}, \cdots, k_n) - F(\infty, \cdots, \infty, k_{n-i+1}, k_{n-i+2} - 1, \cdots, k_n - 1)}{F(\infty, \cdots, \infty, k_{n-i+2}, \cdots, k_n) - F(\infty, \cdots, \infty, k_{n-i+2} - 1, \cdots, k_n - 1)},$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$

$$\frac{F(k_1 - 1, k_2, \cdots, k_n) - F(k_1 - 1, \cdots, k_n - 1)}{F(\infty, k_2, \cdots, k_n) - F(\infty, k_2 - 1, \cdots, k_n - 1)}$$
$$\leqq y_n < \frac{F(k_1, \cdots, k_n) - F(k_1, k_2 - 1, \cdots, k_n - 1)}{F(\infty, k_2, \cdots, k_n) - F(\infty, k_2 - 1, \cdots, k_n - 1)}.$$

By this means we have quasi-rectangular subcells of the unit cell in correspondence with quasi-rectangular subcells of the whole space. Continuing this process indefinitely by further subdivisions, as was done with the 2-dimensional variate, we are able to set up a one-to-one correspondence between the points of the unit cell and the points of the whole n-dimensional space, thereby determining a transformation T carrying over the points of the unit cell to points of the whole space. And hence corresponding to the points of a set D in the unit cell Δ: $0 \leqq y_i < 1$ $(i = 1, 2, \cdots, n)$, which is admissibly ordered with respect to $\Pi(E)$, we have by means of T an n-dimensional variate

$$(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n) = (x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, \cdots, x_n^{(1)}), (x_1^{(2)}, x_2^{(2)}, \cdots, x_n^{(2)}), \cdots$$

which is admissible with respect to the function $F(s_1, s_2, \cdots, s_n)$. Therefore we have

THEOREM 1. *Given an n-dimensional accumulative probability function* $F(s_1, s_2, \cdots, s_n)$, *there exists an n-dimensional variate* $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ *which is admissible with respect to the given function.*

An interesting property of n-dimensional admissible variates is to be found in the following

THEOREM 2. *If* $(\mathbf{x}_{11}, \mathbf{x}_{12}, \cdots, \mathbf{x}_{1n}), (\mathbf{x}_{21}, \mathbf{x}_{22}, \cdots, \mathbf{x}_{2n}), \cdots, (\mathbf{x}_{k1}, \mathbf{x}_{k2}, \cdots, \mathbf{x}_{kn})$ *is an independent set of n-dimensional variates such that* $(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \cdots, \mathbf{x}_{in})$ *is admissible with respect to* $F_i(s_1, s_2, \cdots, s_n)$ *and if* $g_i(s_1, s_2, \cdots, s_n)$ $(i = 1, 2, \cdots, k)$ *is any set of continuous real functions in* $(s_1, s_2, \cdots, s_n)$, *then*

$$\mathbf{p}\left[\prod_{i=1}^{k} g_i(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\phi_{I_i}(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\right] = \prod_{i=1}^{k}\int_{I_i} g_i(s_1, \cdots, s_n)d_nF_i(s_1 + 0, \cdots, s_n + 0)$$

for every set of finite cells $I_1, I_2, \cdots, I_k$, *each of which has the property* I_i: $a_{ij} < s_j \leqq b_{ij}$ $(i = 1, 2, \cdots, k; j = 1, 2, \cdots, n)$.

First we shall prove the theorem for the case when $g_i(s_1, s_2, \cdots, s_n)$ are real and non-negative. Let each cell be subdivided as follows: $a_{ij} = a_{ij}^0 < a_{ij}^1 < \cdots < a_{ij}^{r_i} = b_{ij}$; let I_{it} be the subcell of I_i: $a_{ij}^{t-1} < s_j \leqq a_{ij}^t$; and let m_{it} and M_{it} be respectively the minimum and maximum of $g_i(s_1, s_2, \cdots, s_n)$ in I_{it}. Then

$$\mathbf{p}\left[\prod_{i=1}^{k}\left\{\sum_{t=1}^{r_i} m_{it}\phi_{I_{it}}(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\right\}\right]$$

exists and is less than or, at most, equal to

$$\underline{\mathbf{p}}\left[\prod_{i=1}^{k} g_i(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\phi_{I_i}(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\right].$$

Thus we have

$$\prod_{i=1}^{k}\sum_{t=1}^{r_i} m_{it}\left[\Delta_{a_{i1}^{t-1}, \cdots, a_{in}^{t-1}}^{a_{i1}^t, \cdots, a_{in}^t} F_i(s_1 + 0, \cdots, s_n + 0)\right] \leqq \underline{\mathbf{p}}\left[\prod_{i=1}^{k} g_i(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\phi_{I_i}(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\right].$$

Similarly,

$$\prod_{i=1}^{k}\sum_{t=1}^{r_i} M_{it}\left[\Delta_{a_{i1}^{t-1}, \cdots, a_{in}^{t-1}}^{a_{i1}^t, \cdots, a_{in}^t} F_i(s_1 + 0, \cdots, s_n + 0)\right] \geqq \underline{\mathbf{p}}\left[\prod_{i=1}^{k} g_i(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\phi_{I_i}(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\right].$$

And hence

$$\underline{\mathbf{p}}\left[\prod_{i=1}^{k} g_i(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\phi_{I_i}(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\right] = \prod_{i=1}^{k}\int_{I_i} g_i(s_1, \cdots, s_n)d_nF_i(s_1 + 0, \cdots, s_n + 0).$$

In like manner we may obtain

$$\overline{\mathbf{p}}\left[\prod_{i=1}^{k} g_i(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\phi_{I_i}(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\right] = \prod_{i=1}^{k}\int_{I_i} g_i(s_1, \cdots, s_n)d_nF_i(s_1 + 0, \cdots, s_n + 0).$$

Therefore

$$\mathbf{p}\left[\prod_{i=1}^{k} g_i(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\phi_{I_i}(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in})\right] = \prod_{i=1}^{k}\int_{I_i} g_i(s_1, \cdots, s_n)d_nF_i(s_1 + 0, \cdots, s_n + 0).$$

Now if the functions $g_i(s_1, \cdots, s_n)$ are arbitrary real functions, we may let

$$y_i(s_1, \cdots, s_n) = \begin{cases} g_i(s_1, \cdots, s_n) & \text{if } g_i(s_1, \cdots, s_n) \geqq 0, \\ 0 & \text{if } g_i(s_1, \cdots, s_n) < 0, \end{cases}$$

and

$$\delta_i(s_1, \cdots, s_n) = \begin{cases} 0 & \text{if } g_i(s_1, \cdots, s_n) \geqq 0, \\ -g_i(s_1, \cdots, s_n) & \text{if } g_i(s_1, \cdots, s_n) < 0. \end{cases}$$

Then

$$g_i(s_1, \cdots, s_n) = y_i(s_1, \cdots, s_n) - \delta_i(s_1, \cdots, s_n).$$

Thus this case reduces to the one already treated, with the aid of the additive property of the integral and of the operator $\mathbf{p}(\quad)$.

Another interesting property of admissible n-dimensional variates is given in the following

THEOREM 3. *If $g(s_1, \cdots, s_n, t_1, \cdots, t_n)$ is a continuous function of $(s_1, \cdots, s_n, t_1, \cdots, t_n)$ and is bounded throughout the $2n$-dimensional space, and if $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ is an admissible n-dimensional variate, then $\mathbf{p}_k[g(\mathbf{x}_1, \cdots, \mathbf{x}_n, t_1, \cdots, t_n)]$ converges to its limit $\mathbf{p}[g(\mathbf{x}_1, \cdots, \mathbf{x}_n, t_1, \cdots, t_n)]$ uniformly throughout any finite n-dimensional cell J: $\alpha_i \leqq t_i \leqq \beta_i$ $(i = 1, 2, \cdots, n)$.*

Let B be the least upper bound of the function $|g(s_1, \cdots, s_n, t_1, \cdots, t_n)|$ and let $F(s_1, \cdots, s_n)$ be the accumulative probability function of the n-dimensional variate $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$.

Then, for every $\epsilon > 0$ we can determine a set of numbers a_i and another set b_i such that

$$B \cdot [1 - \Delta_{a_1, \cdots, a_n}^{b_1, \cdots, b_n} F(s_1 + 0, \cdots, s_n + 0)] < \epsilon.$$

Also let I be the cell $a_i < s_i \leqq b_i$ $(i = 1, 2, \cdots, n)$ and CI be the complement of I with respect to the cell $-\infty \leqq s_i \leqq +\infty$ $(i = 1, 2, \cdots, n)$, then we can choose a number k_1 such that

$$| \mathbf{p}_k[g(\mathbf{x}_1, \cdots, \mathbf{x}_n, t_1, \cdots, t_n) \phi_{CI}(\mathbf{x}_1, \cdots, \mathbf{x}_n)] | < \epsilon$$

for every $(t_1, t_2, \cdots, t_n)$, provided $k \geqq k_1$. Hence we have

$$\begin{aligned} | \mathbf{p}_k[g(\mathbf{x}_1, \cdots, \mathbf{x}_n, t_1, \cdots, t_n) \phi_{CI}(\mathbf{x}_1, \cdots, \mathbf{x}_n)] &- \\ \mathbf{p}[g(\mathbf{x}_1, \cdots, \mathbf{x}_n, t_1, \cdots, t_n) \phi_{CI}(\mathbf{x}_1, \cdots, \mathbf{x}_n)] | &< 2\epsilon, \end{aligned}$$

provided $k \geqq k_1$.

Now let I_j be the n-dimensional cell $s_{i,j-1} < s_i \leqq s_{ij}$ $(i = 1, 2, \cdots, n; j = 1, 2, \cdots, r)$, where $s_{i0} = a_i$ and $s_{ir} = b_i$, and let $M_j(t_1, \cdots, t_n)$ and $m_j(t_1, \cdots, t_n)$ be respectively the maximum and the minimum of the function $g(s_1, \cdots, s_n, t_1, \cdots, t_n)$ in the cell I_j. We may choose the cells I_j so that

$$M_j(t_1, \cdots, t_n) - m_j(t_1, \cdots, t_n) < \epsilon$$

for $j = 1, 2, \cdots, r$ for every $(t_1, \cdots, t_n)$ in J. Then

$$\left| \mathbf{p}_k[g(\mathbf{x}_1, \cdots, \mathbf{x}_n, t_1, \cdots, t_n)\phi_I(\mathbf{x}_1, \cdots, \mathbf{x}_n)] - \sum_{j=1}^{r} m_j(t_1, \cdots, t_n)\mathbf{p}_k[\phi_{I_j}(\mathbf{x}_1, \cdots, \mathbf{x}_n)] \right| < \epsilon$$

if $(t_1, \cdots, t_n)$ is in J. Also there exists a number k_2 such that

$$\left| \sum_{j=1}^{r} m_j(t_1, \cdots, t_n)\mathbf{p}_k[\phi_{I_j}(\mathbf{x}_1, \cdots, \mathbf{x}_n)] - \mathbf{p}[g(\mathbf{x}_1, \cdots, \mathbf{x}_n, t_1, \cdots, t_n)\phi_I(\mathbf{x}_1, \cdots, \mathbf{x}_n)] \right| < \epsilon,$$

provided $(t_1, \cdots, t_n)$ is in J and $k \geqq k_2$. Hence

$$| \mathbf{p}_k[g(\mathbf{x}_1, \cdots, \mathbf{x}_n, t_1, \cdots, t_n) \phi_I(\mathbf{x}_1, \cdots, \mathbf{x}_n)] - \mathbf{p}[g(\mathbf{x}_1, \cdots, \mathbf{x}_n, t_1, \cdots, t_n) \phi_I(\mathbf{x}_1, \cdots, \mathbf{x}_n)] | < 2\epsilon,$$

provided $(t_1, \cdots, t_n)$ is in J and $k \geqq k_2$. Therefore

$$| \mathbf{p}_k[g(\mathbf{x}_1, \cdots, \mathbf{x}_n, t_1, \cdots, t_n)] - \mathbf{p}[g(\mathbf{x}_1, \cdots, \mathbf{x}_n, t_1, \cdots, t_n)] | < 4\epsilon,$$

provided $(t_1, \cdots, t_n)$ is in J and $k \geqq K$, where $K = \max [k_1, k_2]$.

Thus $\mathbf{p}_k[g(\mathbf{x}_1, \cdots, \mathbf{x}_n, t_1, \cdots, t_n)]$ approaches its value in the limit uniformly throughout the finite cell J: $\alpha_i \leqq t_i \leqq \beta_i$ $(i = 1, 2, \cdots, n)$. And therefore the theorem is proved.

3. **The probability theorem for n dimensions.** In an article entitled *The probability limit theorem,*[7] A. H. Copeland proved that the assumption that physical measurements behave in accordance with the probability limit theorem does not imply any inconsistency in the mathematical sense. This theorem, however, was for one-dimensional variates. It is our purpose, therefore, to extend his results so as to apply to variates in n dimensions. Hence, his theorem may be viewed as a particular case of the theorem to be established in this section. The generalized theorem reads as follows:

THEOREM 4. *Let* $(\mathbf{x}_{11}, \mathbf{x}_{12}, \cdots, \mathbf{x}_{1n})$, $(\mathbf{x}_{21}, \mathbf{x}_{22}, \cdots, \mathbf{x}_{2n})$, $\cdots$, $(\mathbf{x}_{k1}, \mathbf{x}_{k2}, \cdots, \mathbf{x}_{kn})$, $\cdots$ *be an infinite independent set of independent n-dimensional variates such that* $\mathbf{p}(\mathbf{x}_{ki}) = 0$, *where* $i = 1, 2, \cdots, n$; $k = 1, 2, \cdots$. *And let* $(\mathbf{X}_{m1}, \mathbf{X}_{m2}, \cdots, \mathbf{X}_{mn})$ *be the n-dimensional variate defined by*

$$\mathbf{X}_{mi} = \frac{\mathbf{x}_{1i} + \mathbf{x}_{2i} + \cdots + \mathbf{x}_{mi}}{B_{mi}},$$

where

$$B_{mi} = b_{1i}^2 + b_{2i}^2 + \cdots + b_{mi}^2 \quad \text{and} \quad b_{ki}^2 = \mathbf{p}(\mathbf{x}_{ki}^2).$$

If, given any set of positive numbers $\epsilon_i > 0$, *there exist a set of numbers* a_i *and a number* M *such that*

$$\mathbf{p}[\phi_{|s_i| \geq a_i}(\mathbf{x}_{ki}) \cdot \mathbf{x}_{ki}^2] < b_{ki}^2 \cdot \epsilon_i$$

for every k *and* i, *and such that* $b_{ki}^2/B_{mi}^2 < \epsilon_i$ *for every* $k \leqq m$ *whenever* $m \geqq M$, *then*

$$\lim_{m \to \infty} \mathbf{p}[\phi_{I_s}(\mathbf{X}_{m1}, \cdots, \mathbf{X}_{mn})] = \frac{1}{(2\pi)^{\frac{1}{2}n}} \int_{-\infty}^{s_1} \int_{-\infty}^{s_2} \cdots \int_{-\infty}^{s_n} \exp\left[-\frac{1}{2}\sum_1^n t_i^2\right] dt_1 \cdots dt_n,$$

where I_s *is the n-dimensional cell defined by* $-\infty \leqq t_i \leqq s_i$ $(i = 1, 2, \cdots, n)$.

Before proceeding to prove the theorem,[8] we will establish the three following lemmas.

LEMMA 1. *Under the hypotheses of Theorem* 4

$$\lim_{m \to \infty} \mathbf{p}\left[\exp\left(i \sum_1^n \mathbf{X}_{mj} t_j\right)\right] = \exp\left[-\frac{1}{2}\sum_1^n t_j^2\right].$$

This follows directly from Lemma 1 of Copeland's paper[9] by the application of a well-known theorem on limits.

[7] This Journal, vol. 2(1936), pp. 171–176.

[8] An alternative proof which is based on a well-known theorem on limits suggested itself to the author. But inasmuch as it would eliminate a direct parallel to Copeland's work as well as a simplification of one of Copeland's theorems, the method used was thought to be more pertinent to the problem under discussion.

[9] This Journal, vol. 2(1936), p. 173.

LEMMA 2.[10] *If* $(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ *is an independent n-dimensional variate whose distribution function* $F(s_1, \cdots, s_n)$ *exists and if*

$$F(s_1, \cdots, s_n, h_1, \cdots, h_n) = \frac{1}{(2\pi i)^n} \int_{-h_1}^{+h_1} \int_{-h_2}^{+h_2} \cdots \int_{-h_n}^{+h_n} \prod_{k=1}^{n} \left(\frac{e^{it_k} - \mathbf{p}(e^{i\mathbf{x}_k t_k}) e^{-is_k t_k}}{t_k} \right) dt_1 \cdots dt_n,$$

then

$$F(s_1 - \epsilon, \cdots, s_n - \epsilon) - \epsilon < F(s_1, \cdots, s_n, h_1, \cdots, h_n) < F(s_1 + \epsilon, \cdots, s_n + \epsilon) + \epsilon$$

for every $h_k \geqq h_k'$ $(k = 1, 2, \cdots, n)$ *dependent upon the choice of* ϵ.

The function

$$\prod_{k=1}^{n} \left(\frac{e^{it_k} - e^{i\mathbf{x}_k t_k} e^{-is_k t_k}}{t_k} \right)$$

is bounded and continuous in $(s_1, \cdots, s_n, h_1, \cdots, h_n)$ throughout the $2n$-dimensional space. Thus as $m \to \infty$

$$\mathbf{p}_m \left[\prod_{k=1}^{n} \left(\frac{e^{it_k} - e^{i\mathbf{x}_k t_k} e^{-is_k t_k}}{t_k} \right) \right]$$

converges uniformly[11] and hence

$$\mathbf{p}\left[\frac{1}{(2\pi i)^n} \int_{-h_1}^{+h_1} \cdots \int_{-h_n}^{+h_n} \prod_{k=1}^{n} \left(\frac{e^{it_k} - e^{i\mathbf{x}_k t_k} e^{-is_k t_k}}{t_k} \right) dt_1 \cdots dt_n \right] = F(s_1, \cdots, s_n, h_1, \cdots, h_n).$$

[10] It is interesting to note that Copeland's Lemma 2 (ibid., p. 174) may be modified to read as follows:

If $\mathbf{x}$ *is a variate whose distribution function* $F(s)$ *exists, and if*

$$F(s, h) = \frac{1}{2\pi i} \int_{-h}^{+h} \frac{e^{it} - \mathbf{p}(e^{i\mathbf{x}t}) e^{-ist}}{t} \, dt,$$

then $F(s - \epsilon) - \epsilon < F(s, h) < F(s + \epsilon) + \epsilon$ *for every* $h \geqq h_1$ *dependent upon the choice of* ϵ.

The proof of this is briefly as follows: From Copeland's original lemma we have

$$F(s) - \tfrac{1}{2}[F(s + \epsilon) - F(s - \epsilon)] - 2\epsilon < F(s, h) < F(s) + \tfrac{1}{2}[F(s + \epsilon) - F(s - \epsilon)] + 2\epsilon.$$

By taking the limit we obtain

$$F(s - 0) \leq \underline{\lim}_{h \to \infty} F(s, h) \leq \overline{\lim}_{h \to \infty} F(s, h) \leq F(s + 0)$$

since by (2) (Copeland, ibid., p. 173)

$$F(s) = F(s - 0) + \frac{F(s + 0) - F(s - 0)}{2}.$$

Therefore for every $h \geqq h_1$ (dependent upon the choice of ϵ) $F(s - \epsilon) - \epsilon < F(s, h) < F(s + \epsilon) + \epsilon$. This modified lemma enables us to simplify considerably the proof of the ultimate result obtained by Copeland. The revised proof follows directly the method used in the present paper.

[11] See Theorem 3 above.

Now if

$$F_k(s_k, h_k) = \mathbf{p}\left[\frac{1}{2\pi i}\int_{-h_k}^{+h_k} \frac{e^{it_k} - e^{i\mathbf{X}_k t_k}e^{-is_k t_k}}{t_k}\, dt_k\right],$$

then

$$F(s_1, \cdots, s_n, h_1, \cdots, h_n) = \prod_{k=1}^{n} F_k(s_k, h_k),$$

and by hypothesis

$$F(s_1, \cdots, s_n) = \prod_{k=1}^{n} F_k(s_k).$$

But by the modified lemma, given above in footnote 10, we have

$$F_k(s_k - \epsilon) - \epsilon < F_k(s_k, h_k) < F_k(s_k + \epsilon) + \epsilon,$$

and consequently

$$F(s_1 - \epsilon, \cdots, s_n - \epsilon) - \epsilon_1 < F(s_1, \cdots, s_n, h_1, \cdots, h_n) < F(s_1 + \epsilon, \cdots, s_n + \epsilon) + \epsilon.$$

LEMMA 3. *If* $\Phi(s_1, \cdots, s_n) = \dfrac{1}{(2\pi)^{\frac{1}{2}n}} \displaystyle\int_{-\infty}^{s_1} \cdots \int_{-\infty}^{s_n} \exp\left[-\frac{1}{2}\sum_{k=1}^{n} t_k^2\right] dt_1 \cdots dt_n,$ *and if*

$$\Phi(s_1, \cdots, s_n, h_1, \cdots, h_n) = \frac{1}{(2\pi i)^n}\int_{-h_1}^{+h_1} \cdots \int_{-h_n}^{+h_n} \prod_{k=1}^{n}\left(\frac{e^{it_k} - e^{-\frac{1}{2}t_k^2}e^{-is_k t_k}}{t_k}\right) dt_1 \cdots dt_n,$$

then

$$\Phi(s_1, \cdots, s_n) - \epsilon < \Phi(s_1, \cdots, s_n, h_1, \cdots, h_n) < \Phi(s_1, \cdots, s_n) + \epsilon.$$

This follows immediately from Copeland's Lemma 3 in the same way as given above in the proof of our Lemma 2.

Now if we let $\Phi_m(s_1, \cdots, s_n) = \mathbf{p}[\phi_{I_s}(\mathbf{X}_{m1}, \cdots, \mathbf{X}_{mn})]$, and if

$$\Phi_m(s_1, \cdots, s_n, h_1, \cdots, h_n) = \frac{1}{(2\pi i)^n}\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{k=1}^{n}\left(\frac{e^{it_k} - \mathbf{p}(e^{i\mathbf{X}_{mk}t_k})e^{-is_k t_k}}{t_k}\right) dt_1 \cdots dt_n,$$

then from Lemma 1 we have

$$\Phi(s_1, \cdots, s_n, h_1, \cdots, h_n) - \epsilon < \Phi_m(s_1, \cdots, s_n, h_1, \cdots, h_n) < \Phi(s_1, \cdots, s_n, h_1, \cdots, h_n) + \epsilon \tag{3.1}$$

for every $m \geqq M$; and by Lemma 2

$$\Phi_m(s_1 - \epsilon, \cdots, s_n - \epsilon) - \epsilon < \Phi_m(s_1, \cdots, s_n, h_1, \cdots, h_n) < \Phi_m(s_1 + \epsilon, \cdots, s_n + \epsilon) + \epsilon \tag{3.2}$$

and by Lemma 3

$$(3.3)\quad \Phi(s_1, \cdots, s_n) - \epsilon < \Phi(s_1, \cdots, s_n, h_1, \cdots, h_n) < \Phi(s_1, \cdots, s_n) + \epsilon.$$

Combining (3.1), (3.2) and (3.3) we obtain

$$(3.4)\qquad \Phi_m(s_1 - \epsilon, \cdots, s_n - \epsilon) < \Phi(s_1, \cdots, s_n) + 3\epsilon.$$

Substituting $s_j + \epsilon$ for s_j in (3.4) ($j = 1, 2, \cdots, n$), we have

$$\Phi_m(s_1, \cdots, s_n) < \Phi(s_1, \cdots, s_n) + \epsilon',$$

since

$$\Phi(s_1 + \epsilon, \cdots, s_n + \epsilon) - \Phi(s_1, \cdots, s_n) < \epsilon \cdot k.$$

Similarly, it may be shown that

$$\Phi_m(s_1, \cdots, s_n) > \Phi(s_1, \cdots, s_n) - \epsilon'.$$

Therefore

$$\lim_{m \to \infty} \Phi_m(s_1, \cdots, s_n) = \Phi(s_1, \cdots, s_n),$$

and the proof of the theorem is complete.

SAINT LOUIS UNIVERSITY AND SPRING HILL COLLEGE.

LIMIT POINTS OF SEQUENCES AND THEIR TRANSFORMS BY METHODS OF SUMMABILITY

By Henry G. Barone

1. **Introduction.** Let $\{s_n\}$ be a complex sequence, and let A be its set of limit points. The problem which we consider is that of determining, for each of several methods of summability, whether the set of limit points of the transform of each bounded sequence $\{s_n\}$ is a connected set.

In §2 examples are given which show that if no restrictions are placed on the complex sequence $\{s_n\}$, then the set A of its limit points need not be connected. In §3 sufficient conditions that A be connected are given. In §4 theorems concerning transforms of bounded complex sequences and their sets of limit points are proved. We devote §§5, 6, 7, 8, and 9 to the Hölder, Cesàro, Riesz, de la Vallée Poussin, and Euler transforms, respectively, of bounded complex sequences $\{s_n\}$. In each case we determine whether the set of limit points of the transform of $\{s_n\}$ is connected.

2. **Some examples.** Since A is a closed set, to say that it is connected means that A cannot be broken up into two mutually exclusive subsets A_1 and A_2, both closed, unless either $A_1 = \Lambda$ (empty set) or $A_2 = \Lambda$.

To show that A need not be connected, let us consider the following three complex sequences:

$$1, i, 1, i, 1, i, \cdots, \tag{2.1}$$

$$1, i, 2, 1, i, 3, 1, i, 4, \cdots, \tag{2.2}$$

$$\begin{gathered} 0, \frac{i}{10}, \frac{2i}{10}, \cdots, i, \frac{1}{10} + i, \frac{2}{10} + i, \cdots, 1 + i, 1 + \frac{9i}{10}, 1 + \frac{8i}{10}, \cdots, 1, \\ 1 + \frac{i}{20}, 1 + \frac{2i}{20}, \cdots, 1 + 2i, \frac{19}{20} + 2i, \frac{18}{20} + 2i, \cdots, 2i, \frac{39i}{20}, \cdots, 0, \\ \frac{i}{30}, \frac{2i}{30}, \cdots, 3i, \frac{1}{30} + 3i, \frac{2}{30} + 3i, \\ \cdots, 1 + 3i, 1 + \frac{89i}{30}, 1 + \frac{88i}{30}, \cdots, 1, \\ 1 + \frac{i}{40}, 1 + \frac{2i}{40}, \cdots, 1 + 4i, \frac{39}{40} + 4i, \frac{38}{40} + 4i, \\ \cdots, 4i, \frac{159i}{40}, \frac{158i}{40}, \cdots, 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots. \end{gathered} \tag{2.3}$$

Received November 8, 1938.

For sequences (2.1) and (2.2), $A = 1, i$; and for (2.3) A consists of all those points not below the real axis with real part 0 or 1. These sets are not connected.

Sequence (2.1) is bounded, while (2.2) and (2.3) are unbounded. Sequence (2.3) has the following property:

$$\lim_{n\to\infty} |s_{n+1} - s_n| = 0, \tag{2.4}$$

while neither (2.1) nor (2.2) has this property.

3. **Sufficient conditions that A be connected.** We now prove the following theorem.

THEOREM 3.1. *If $\{s_n\}$ is a bounded complex sequence such that*

$$\lim_{n\to\infty} |s_{n+1} - s_n| = 0,$$

then the set of its limit points is connected.

This is a special case of the following theorem.

THEOREM 3.2. *If $\{s_n\}$ is a compact sequence in a metric space such that*

$$\lim_{n\to\infty} r(s_n, s_{n+1}) = 0, \tag{3.21}$$

where $r(s_n, s_{n+1})$ denotes the distance between the elements s_n and s_{n+1} of the metric space, then the set A of its limit points is connected.[1]

Let us suppose that A is not connected. Then A can be broken up into two mutually exclusive subsets A_1 and A_2 such that neither $A_1 = \Lambda$ nor $A_2 = \Lambda$.

Since $A_1A_2 = \Lambda$, and A_1 and A_2 are closed and compact, $r(A_1, A_2) = \rho > 0$. Also there exist $a_1 \in A_1$, $a_2 \in A_2$ such that $r(a_1, a_2) = \rho$.

Since $a_1, a_2 \in A$, there exist subsequences of $\{s_n\}$,

$$s_{k_1}, s_{k_2}, s_{k_3}, \cdots, \tag{3.22}$$

$$s_{l_1}, s_{l_2}, s_{l_3}, \cdots, \tag{3.23}$$

such that $\lim_{n\to\infty} s_{k_n} = a_1$, $\lim_{n\to\infty} s_{l_n} = a_2$, and $k_1 < l_1 < k_2 < l_2 < \cdots$. For a given $\epsilon > 0$ and the above ρ, there exists an N such that

$$r(s_{k_n}, a_1) < \tfrac{1}{4}\rho, \qquad r(s_{l_n}, a_2) < \tfrac{1}{4}\rho, \qquad r(s_m, s_{m+1}) < \tfrac{1}{4}\rho \tag{3.24}$$

for all $k_n, l_n, m > N$.

For each $k_n > N$ let us consider the group

$$s_{k_n}, s_{k_n+1}, s_{k_n+2}, \cdots, s_{l_n} \tag{3.25}$$

[1] The terminology and theorems used in the proof of this theorem are those given in H. Hahn, *Theorie der Reellen Funktionen*, 1921, Chapter 1.

of elements of $\{s_n\}$. Since (3.24) holds, we have for $m, k_n > N$

$$r(s_{k_n}, A_1) < \tfrac{1}{4}\rho, \qquad r(s_{l_n}, A_2) < \tfrac{1}{4}\rho, \qquad r(s_m, s_{m+1}) < \tfrac{1}{4}\rho. \tag{3.26}$$

Hence there must exist some index p_n such that $k_n < k_n + p_n < l_n$ and

$$r(s_{k_n+p_n}, A_1) > \tfrac{1}{4}\rho, \qquad r(s_{k_n+p_n}, A_2) > \tfrac{1}{4}\rho, \qquad k_n > N. \tag{3.27}$$

This would mean that, for some elements, $r(s_{k_n+p_n}, s_{k_n+p_n+1}) > \frac{1}{2}\rho$, and this is a contradiction.

We now have an infinite sequence

$$s_{k_1+p_1}, s_{k_2+p_2}, s_{k_3+p_3}, \cdots, \tag{3.28}$$

which is a subsequence of $\{s_n\}$ and is such that (3.27) holds. Since $\{s_n\}$ is compact, this subsequence has a limit point, say c, such that

$$r(c, A_1) \geqq \tfrac{1}{4}\rho, \qquad r(c, A_2) \geqq \tfrac{1}{4}\rho. \tag{3.29}$$

Thus $c \in A$, but does not belong to A_1 or A_2 and therefore $A \neq A_1 + A_2$. This is a contradiction. Hence A is connected.

The set of limit points of the sequence $\{s_n\}$ where

$$s_n = e^{in} \qquad (n = 0, 1, 2, 3, \cdots) \tag{3.30}$$

is connected, but condition (2.4) does not hold. Hence (2.4) is not a necessary condition that A be connected.

4. **Theorems concerning transforms and their sets of limit points.** Given a matrix a_{nk} of complex numbers, each complex sequence $\{s_n\}$ has a transform

$$A: \sigma_n = \sum_{k=0}^{n} a_{nk} s_k \qquad (n = 0, 1, 2, \cdots). \tag{4.01}$$

The transformation A is said to be regular if it transforms each convergent sequence into a convergent sequence having the same limit. Also by the Silverman-Toeplitz Theorem, necessary and sufficient conditions that the transformation A be regular are

$$\sum_{k=0}^{n} |a_{nk}| < M \qquad \text{for all } n, \tag{4.02}$$

$$\lim_{n\to\infty} \sum_{k=0}^{n} a_{nk} = 1, \tag{4.03}$$

$$\lim_{n\to\infty} a_{nk} = 0 \qquad \text{for each } k. \tag{4.04}$$

Let us consider the following conditions on the transformations A and B.

$$\sum_{k=0}^{n} |a_{nk}| < M \qquad \text{for all } n, \tag{4.11}$$

$$\lim_{n\to\infty} \sum_{k=0}^{n} |a_{nk} - a_{n-1,k}| = 0 \qquad (a_{n-1,n} = 0), \tag{4.12}$$

(4.13) $$\lim_{n \to \infty} \sum_{k=0}^{n} | a_{nk} - b_{nk} | = 0,$$

(4.14) $$| a_{nn} | - \sum_{k=0}^{n-1} | a_{nk} | \geqq P > 0 \qquad (n = 0, 1, 2, \cdots),$$

(4.15) $$| a_{nn} | > P > 0 \qquad \text{for all } n,$$

(4.16) $$| a_{nk} | \Big/ \sum_{p=k+1}^{n} | a_{np} | \to \infty \qquad \text{as } n \to \infty \quad (k = 0, 1, 2, \cdots),$$

(4.17) $$a_{nk} = f_n, \qquad | a_{nk} | > P > 0,$$

(4.18) $$a_{nk} - r a_{n-1,k} = f_k \qquad \text{for } k < n,$$

(4.19) $$| f_{n-1} - a_{n-1,n-1} | < \rho | a_{nn} |.$$

We shall prove the following theorems, the first three of which follow almost immediately.

THEOREM 4.2. *If the transformation A is such that* (4.11) *holds, then the A-transform of each bounded sequence $\{s_n\}$ is also bounded.*

THEOREM 4.3. *If the transformation A satisfies* (4.11) *and* (4.12), *then the set of limit points of the A-transform of each bounded sequence $\{s_n\}$ is a connected set.*

THEOREM 4.4. *If $\{s_n\}$ is a bounded sequence and the transformations A and B are such that* (4.13) *holds, and if the set of limit points of the A-transform of $\{s_n\}$ is connected, then the set of limit points of the B-transform of $\{s_n\}$ is also connected.*

THEOREM 4.5. *If the transformation A is such that* (4.14) *holds, then the set of limit points of the A-transform of a bounded sequence $\{s_n\}$ need not be connected.*

It is sufficient to show that a sequence $\{\sigma_n\}$ having two limit points is the transform of a bounded sequence $\{s_n\}$. This may be done by showing that if $\{\sigma_n\}$ is bounded, then the sequence $\{s_n\}$ obtained by the inverse transformation is also bounded. This, in turn, may be done by showing that if $\{s_n\}$ is unbounded, then $\{\sigma_n\}$ is unbounded.

For a given $M > 0$ there exists an n such that

(4.51) $$| s_n | > M/P, \qquad | s_k | < | s_n | \text{ for } k < n.$$

Then we have

(4.52) $$\begin{aligned} | \sigma_n | &\geqq - \left| \sum_{k=0}^{n-1} a_{nk} s_k \right| + | a_{nn} | \cdot | s_n | \\ &\geqq - | s_n | \cdot \sum_{k=0}^{n-1} | a_{nk} | + | a_{nn} | \cdot | s_n | \\ &\geqq | s_n | \cdot \left\{ | a_{nn} | - \sum_{k=0}^{n-1} | a_{nk} | \right\} \geqq M, \end{aligned}$$

and the theorem is proved.

THEOREM 4.6. *If the transformation A is such that* (4.15) *and* (4.16) *hold, then the set of limit points of the A-transform of a bounded sequence $\{s_n\}$ is connected.*

If $s_n = 0$ for all n, then $\sigma_n = 0$ for all n. If $s_n \neq 0$ for some n, $| s_n | < M$ for all n, and if k is the first index such that $s_k \neq 0$, then we have

$$(4.61)\qquad \begin{aligned} | \sigma_n | &\geqq | a_{nk} | \cdot | s_k | - M \sum_{p=k+1}^{n} | a_{np} | \\ &= \sum_{p=k+1}^{n} | a_{np} | \left\{ | a_{nk} | \cdot | s_k | \Big/ \sum_{p=k+1}^{n} | a_{np} | - M \right\}, \end{aligned}$$

and this becomes infinite as $n \to \infty$. Hence the theorem follows.

THEOREM 4.7. *If the transformation A is such that* (4.17) *holds, then the set of limit points of the A-transform of a bounded sequence $\{s_n\}$ need not be connected.*

The inverse transformation is given by

$$(4.71)\qquad s_n = \frac{\sigma_n}{f_n} - \frac{\sigma_{n-1}}{f_{n-1}},$$

hence $\{s_n\}$ is bounded when $\{\sigma_n\}$ is bounded and so the theorem follows.

THEOREM 4.8. *If the transformation A is such that* (4.15), (4.18), *and* (4.19) *hold, where r is a complex constant and* $0 < \rho < 1$, *then the set of limit points of the A-transform of a bounded sequence $\{s_n\}$ need not be connected.*

This theorem is proved like Theorem 4.4. For a given $M > 0$ there exists an n such that

$$(4.81)\qquad | s_n | > M/P(1 - \rho), \qquad | s_k | < | s_n | \text{ for } k < n.$$

We thus have

$$(4.82)\qquad \begin{aligned} | \sigma_n - (r + 1)\sigma_{n-1} + r\sigma_{n-2} | &= | a_{nn}\{s_n + (f_{n-1} - a_{n-1,n-1})s_{n-1}/a_{nn}\} | \\ &\geqq | a_{nn} | \cdot (| s_n | - | s_{n-1} | \cdot | f_{n-1} - a_{n-1,n-1} |/| a_{nn} |) \\ &\geqq | a_{nn} | \cdot (| s_n | - \rho | s_n |) > M. \end{aligned}$$

Therefore $\{\sigma_n\}$ is unbounded when $\{s_n\}$ is unbounded, and Theorem 4.8 is proved.

5. **The Hölder transformation.** Given a sequence $\{s_n\}$ of complex numbers such that $| s_n | < M$, let us consider its Hölder transform[2]

$$(5.1)\qquad H_n^{(r)} = \{H_0^{(r-1)} + H_1^{(r-1)} + \cdots + H_n^{(r-1)}\}/(n + 1) \qquad (n = 0, 1, 2, \cdots),$$

where r is a positive integer, $H_n^{(0)} = s_n$, and $H_n^{(1)}$ is the arithmetic mean transform of the sequence $\{s_n\}$.

[2] See, for example, K. Knopp, *Theory and Application of Infinite Series*, pp. 465–466.

For each positive integer r, the transformation is regular; hence it follows from Theorem 4.2 that $|H_n^{(r)}| < M$. For a positive integer r and for $n > 0$

$$
\begin{aligned}
|H_n^{(r)} - H_{n-1}^{(r)}| &= \left| \sum_{k=0}^{n} H_k^{(r-1)}/(n+1) - \sum_{k=0}^{n-1} H_k^{(r-1)}/n \right| \\
&= \left| \sum_{k=0}^{n-1} H_k^{(r-1)} \left(\frac{1}{n+1} - \frac{1}{n} \right) + H_n^{(r-1)}/(n+1) \right| \qquad (5.2) \\
&\leqq \sum_{k=0}^{n-1} M/[(n+1)n] + M/(n+1) \leqq 2M/(n+1).
\end{aligned}
$$

We thus have

THEOREM 5.3. *If $\{s_n\}$ is a bounded complex sequence, then the set of limit points of its Hölder transform of order r, where r is any positive integer, is connected.*

6. **The Cesàro transformation.** The Cesàro transform[3] of any order r, real or complex, except a negative real integer, of a sequence $\{s_n\}$ is defined by

$$
C_r:\ \sigma_n = \sum_{k=0}^{n} a_{nk} s_k \qquad (n = 0, 1, 2, \cdots), \tag{6.11}
$$

where

$$
a_{nk} = \binom{r+n-k-1}{n-k} \Big/ \binom{r+n}{n} = \frac{r\Gamma(r+n-k)\Gamma(n+1)}{\Gamma(n-k+1)\Gamma(r+n+1)}. \tag{6.12}
$$

Since the Cesàro transformation is regular when $Re(r) > 0$ or $r = 0$, then, for these values of r, $\{\sigma_n\}$ is bounded when $\{s_n\}$ is bounded.

Let us suppose that $Re(r) > 0$; i.e., let $r = a + ib$, $a > 0$. We wish to show that $\sum_{k=0}^{n} |b_{nk}| \to 0$ as $n \to \infty$, where

$$
b_{nk} = \begin{cases} a_{nk} - a_{n-1,k}, & k < n, \\ a_{nn}, & k = n. \end{cases} \tag{6.13}
$$

For $k < n$

$$
\begin{aligned}
b_{nk} &= \frac{r\Gamma(r+n-k-1)\Gamma(n)}{\Gamma(n-k+1)\Gamma(r+n+1)}(rk - n) \\
&= \left\{ \frac{\Gamma(r+n-k-1)}{\Gamma(n-k+1)\cdot(n-k+1)^{r-2}} \cdot \frac{\Gamma(n)\cdot n^{r+1}}{\Gamma(r+n+1)} \right\} \frac{r(n-k+1)^{r-2}}{n^{r+1}}(rk - n),
\end{aligned} \tag{6.14}
$$

[3] See, for example, E. Kogbetliantz, *Sommation des Séries et Intégrales Divergentes par les Moyennes Arithmétiques et Typiques*, p. 16.

where the expression in braces is uniformly bounded in n and k; and for $k = n$ we have

$$b_{nn} = 1 \Big/ \binom{r+n}{n} = \frac{\Gamma(r+1)\cdot\Gamma(n+1)}{\Gamma(r+n+1)} \tag{6.15}$$
$$= \frac{\Gamma(n+1)(n+1)^r}{\Gamma(r+n+1)} \cdot \frac{\Gamma(r+1)}{(n+1)^r}.$$

Hence we have

$$| b_{nk} | \leqq M(n - k + 1)^{a-2}/n^a \qquad (k = 0, 1, 2, \cdots, n); \tag{6.16}$$

and so

$$S_n \equiv \sum_{k=0}^{n} | b_{nk} | \leqq M \sum_{k=0}^{n} (n - k + 1)^{a-2}/n^a. \tag{6.17}$$

For $a \geqq 2$

$$S_n \leqq M \sum_{k=0}^{n} (n + 1)^{a-2}/n^a \leqq M(n + 1)^{a-1}/n^a. \tag{6.18}$$

For $1 < a < 2$, $(n - k)^{a-2} < 1$, hence

$$S_n \leqq \sum_{k=0}^{n} M/n^a \leqq M(n + 1)/n^a. \tag{6.19}$$

For $a = 1$

$$S_n \leqq \sum_{k=0}^{n} M/[n(n - k + 1)], \tag{6.20}$$

and for $0 < a < 1$

$$S_n \leqq n^{-a} \sum_{k=0}^{n} M/(n - k + 1)^{2-a} \leqq n^{-a} \sum_{k=0}^{\infty} M/(k + 1)^{2-a}, \tag{6.21}$$

where the series converges. In each case $S_n \to 0$ as $n \to \infty$.[4] Hence from Theorem 4.2 we have

THEOREM 6.3. *The set of limit points of the Cesàro transform of each bounded complex sequence is connected when $Re(r) > 0$.*

For $r = 0$ we have the identity transformation and example (2.1) shows that the set of limit points of the transform of a bounded sequence need not be connected.

For $Re(r) \leqq 0$, $r \neq 0$, the Cesàro transformation is not regular since $\sum_{k=0}^{n} | a_{nk} | \to \infty$ as $n \to \infty$. For these values of r we wish to show that the set of limit points of the transform of a bounded sequence need not be connected. This may be done by showing that a sequence $\{\sigma_n\}$ having two limit points may be transformed by the inverse transformation into a bounded sequence $\{s_n\}$.

[4] R. G. Cooke, in Proceedings of the London Mathematical Society, (2), vol. 41(1936), pp. 117–118, proves this result for r real and > 0.

By induction we may show that

$$\binom{r}{n} = \sum_{k=0}^{p} (-1)^k \binom{p}{k} \binom{r+p-k}{n-k}, \tag{6.31}$$

and also

$$\sum_{k=0}^{2^p-1}{}' \pm \binom{r}{n-k} = \sum_{\alpha=0}^{2^p-p-1} (-1)^\alpha A_\alpha \binom{r+2^p-1-\alpha}{n-\alpha}, \tag{6.32}$$

where

$$A_\alpha = \sum_{j_1+j_2+\cdots+j_{p-1}=\alpha} \binom{2}{j_1}\binom{4}{j_2}\cdots\binom{2^{p-1}}{j_{p-1}}, \qquad j_1 < 2,\ j_2 < 2^2,\ \cdots,\ j_{p-1} < 2^{p-1}, \tag{6.33}$$

and $\sum'$ denotes the sum of terms where the first two terms have $+$ signs, the next two have signs opposite to those of the first two, the next four terms have signs opposite to those of the first four, etc.

Let us suppose $r = -a + ib$, $r \neq -a$, $a \geqq 0$, and choose an integer $p > 0$ such that $a < p$. We know that

$$\left|\binom{r+p}{n}\right| = \left|\binom{n-p-r-1}{n}\right| \leqq M/(n+1)^{p+1-a}, \tag{6.34}$$

and hence

$$\left|\binom{r+p+j}{n}\right| \leqq M/(n+1)^{p+j+1-a} \leqq M/(n+1)^{p+1-a} \quad (j = 0, 1, 2, \cdots). \tag{6.35}$$

From (6.32) and (6.35) we see that

$$\left|\sum_{k=\alpha 2^p}^{(\alpha+1)2^p-1}{}' \pm \binom{r}{n-k}\right| \leqq M/\{n-(\alpha+1)2^p+p+2\}^{p+1-a} \quad (\alpha = 0, 1, 2, \cdots). \tag{6.36}$$

From (6.34) and (6.35) it follows that

$$\left|\binom{r+n}{n}\right| \leqq M/(n+1)^a, \tag{6.37}$$

which is bounded for $a \geqq 0$.

For each $\alpha = 0, 1, 2, \cdots$, let σ_n be defined when

$$\alpha \cdot 2^p \leqq n < (\alpha+1)2^p - 1 \tag{6.38}$$

by the formula

$$\sigma_n = \pm A_\alpha \Big/ \binom{n+r}{r} \qquad (n = 0, 1, 2, \cdots), \tag{6.39}$$

where

$$\Lambda_\alpha = \binom{\alpha 2^p + r}{r} \tag{6.40}$$

and the signs of the groups of 2^p terms as given by (6.39) are as follows: the first two terms have signs $+$, $-$; the next two terms have signs opposite to those of the first two; the next four terms have signs opposite to those of the first four; etc.

If we write

$$\sigma_n = \pm\left(1 - \frac{r}{r+n}\right)\left(1 - \frac{r}{r+n-1}\right)\cdots\left(1 - \frac{r}{\alpha 2^p + r + 1}\right), \tag{6.41}$$

it is easily seen that $+1$ and -1 are limit points of $\{\sigma_n\}$.

Substituting σ_k in the inverse Cesàro transformation

$$C_r^{-1}:\ s_n = \sum_{k=0}^{n} (-1)^{n-k}\binom{r}{n-k}\binom{k+r}{r}\sigma_k, \tag{6.42}$$

which may be written

$$s_n = \sum_{\alpha=0}^{\infty}\left\{\sum_{k=\alpha 2^p}^{(\alpha+1)2^p-1} (-1)^{n-k}\binom{r}{n-k}\binom{k+r}{r}\sigma_k\right\}, \tag{6.43}$$

we get

$$s_n = \sum_{\alpha=0}^{\infty}\Lambda_\alpha\left\{\sum_{k=\alpha 2^p}^{(\alpha+1)2^p-1}{}' \pm \binom{r}{n-k}\right\}. \tag{6.44}$$

Applying (6.36) and (6.37) to (6.44), we find

$$\begin{aligned} |s_n| &\leqq \sum_{\alpha=0}^{\infty} |\Lambda_\alpha|\cdot\left|\sum_{k=\alpha 2^p}^{(\alpha+1)2^p-1}{}' \pm \binom{r}{n-k}\right| \\ &\leqq \sum_{\alpha=0}^{\infty} M/(k+1)^{p+1-a}, \end{aligned} \tag{6.45}$$

the last series being convergent since $p > a$. Thus the sequence $\{s_n\}$ is bounded and we have

THEOREM 6.5. *The set of limit points of the Cesàro transform of a bounded sequence need not be connected when $Re(r) \leqq 0$.*

7. **The Riesz transformation.** A series $\sum_{k=0}^{\infty} u_k$ is said to be summable by the method of Riesz[5] if

$$\sigma_n = \sum_{k=0}^{n} d_{nk} u_k, \qquad d_{nk} = \left(1 - \frac{k}{n+1}\right)^r \quad (n = 0, 1, 2, \cdots) \tag{7.01}$$

[5] See M. Riesz, Comptes Rendus, vol. 194(1909), pp. 18-21.

approaches a limit as $n \to \infty$. To convert this to a sequence-to-sequence transformation, let $s_n = u_0 + u_1 + \cdots + u_n$, then

$$a_{nk} = \begin{cases} d_{nk} - d_{n,k+1}, & k < n, \\ d_{nn}, & k = n, \end{cases} \tag{7.02}$$

and from (6.13) and (7.02) we have

$$b_{nk} = \begin{cases} d_{nk} - d_{n,k+1} - d_{n-1,k} + d_{n-1,k+1}, & k < n - 1, \\ d_{n,n-1} - d_{nn} - d_{n-1,n-1}, & k = n - 1, \\ d_{nn}, & k = n. \end{cases} \tag{7.03}$$

Since the Riesz transformation is regular when $Re(r) > 0$ or $r = 0$, and is the identity transformation when $r = 0$, we wish to prove that if $r = a + ib$ with $a > 0$, then $\sum_{k=0}^{n} |b_{nk}| \to 0$ as $n \to \infty$. For $k < n - 1$

$$b_{nk} = \frac{(n - k + 1)^{r-2}}{n^r}\left[A'_{nk} - 2A_{nk} + \frac{A_n A_{nk}}{n + 1} - \frac{rA_n(n - k + 1)}{n + 1}\right], \tag{7.04}$$

where A_n, A_{nk}, A'_{nk} are defined by

$$\left(1 - \frac{1}{n + 1}\right)^r = 1 - \frac{A_n}{n + 1}, \tag{7.05}$$

$$\left(1 - \frac{1}{n - k + 1}\right)^r = 1 - \frac{r}{n - k + 1} + \frac{A_{nk}}{(n - k + 1)^2}, \tag{7.06}$$

$$\left(1 - \frac{2}{n - k + 1}\right)^r = 1 - \frac{2r}{n - k + 1} + \frac{A'_{nk}}{(n - k + 1)^2} \tag{7.07}$$

and are uniformly bounded in n and k. It then follows that for $k < n - 1$

$$|b_{nk}| \leqq M(n - k + 1)^{a-2}/n^a. \tag{7.08}$$

This also holds for $k = n - 1$ and n and so (6.16) holds. Hence $\sum_{k=0}^{n} |b_{nk}| \to 0$ as $n \to \infty$, and we have

THEOREM 7.1. *The set of limit points of the Riesz transform of each bounded complex sequence is connected when $Re(r) > 0$.*

For r real and $\leqq 0$ we have for the Riesz transformation

$$|a_{nk}| = \begin{cases} \{(n - k)^r - (n - k + 1)^r\}/(n + 1)^r, & k < n, \\ 1/(n + 1)^r, & k = n. \end{cases} \tag{7.11}$$

Then

$$\sum_{k=0}^{n-1} |a_{nk}| = 1/(n + 1)^r - 1 \tag{7.12}$$

and

$$(7.13) \qquad | a_{nn} | - \sum_{k=0}^{n-1} | a_{nk} | = 1,$$

hence from Theorem 4.5 follows

THEOREM 7.2. *The set of limit points of the Riesz transform of a bounded sequence need not be connected when r is real and* $\leqq 0$.

8. The de la Vallée Poussin transformation. A series $\sum_{k=0}^{\infty} u_k$ is said to be summable by the method of de la Vallée Poussin[6] to the value s if

$$(8.11) \qquad \sigma_n = \sum_{k=0}^{n} d_{nk} u_k, \qquad d_{nk} = \frac{n!n!}{(n-k)!(n+k)!} \qquad (n = 0, 1, 2, \ldots)$$

approaches s as $n \to \infty$.

Since the de la Vallée Poussin transformation is regular, we wish to prove $\sum_{k=0}^{n} | b_{nk} | \to 0$ as $n \to \infty$. From (7.03) and (8.11) we have

$$(8.12) \qquad b_{nk} = \begin{cases} \dfrac{(n-1)!(n-1)!(2k+1)(k^2+k-n)}{(n-k)!(n+1+k)!}, & k < n-1, \\[2ex] \dfrac{(n-1)!(n-1)!(n-2)}{2(2n-2)!}, & k = n-1, \\[2ex] \dfrac{n!n!}{(2n)!}, & k = n. \end{cases}$$

We notice that for $n > 2$, b_{nk} is real and > 0 when $k > \frac{1}{2}[-1 + (1+4n)^{\frac{1}{2}}]$. If we let $\alpha \equiv \alpha(n)$ be the greatest integer $\leqq \frac{1}{2}[-1 + (1+4n)^{\frac{1}{2}}]$, then

$$(8.13) \qquad b_{nk} \begin{cases} \leqq 0, & k \leqq \alpha, \\ > 0, & k > \alpha. \end{cases}$$

Since we have

$$(8.14) \qquad \sum_{k=0}^{n} b_{nk} = \sum_{k=0}^{n} a_{nk} - \sum_{k=0}^{n-1} a_{n-1,k} = 0,$$

we obtain

$$(8.15) \qquad \begin{aligned} S_n \equiv \sum_{k=0}^{n} | b_{nk} | &= \sum_{k=0}^{\alpha} - b_{nk} + \sum_{k=\alpha+1}^{n} b_{nk} \\ &= 2 \sum_{k=0}^{\alpha} (a_{n-1,k} - a_{nk}). \end{aligned}$$

[6] T. H. Gronwall, Annals of Mathematics, (2), vol. 33(1932), p. 101.

This then simplifies to

$$S_n = 2(d_{n,\alpha+1} - d_{n-1,\alpha+1}) \le \frac{1}{2}\left[\frac{1}{n} + \left(\frac{1}{n^2} + \frac{4}{n}\right)^{\frac{1}{2}}\right]^2. \tag{8.16}$$

Hence, $S_n \to 0$ as $n \to \infty$, and we have

THEOREM 8.2. *The set of limit points of the de la Vallée Poussin transform of each bounded complex sequence is connected.*

9. **The Euler transformation.** The Euler transformation[7] of order r, where r is any complex number, is defined by

$$E_r\colon \sigma_n = \sum_{k=0}^{n} \binom{n}{k} (1-r)^{n-k} r^k s_k \qquad (n = 0, 1, 2, \cdots), \tag{9.11}$$

E_1 being interpreted as the identity.

The Euler transformation is regular for $0 < r \le 1$. For $r = 0$ we have $\sigma_n = s_0$ and $\sigma_n - \sigma_{n-1} = 0$. For $0 < r < 1$ we wish to show that $\sum_{k=0}^{n} | b_{nk} | \to 0$ as $n \to \infty$. We have

$$b_{nk} = \begin{cases} -r(1-r)^{n-1}, & k = 0, \\ \binom{n-1}{k-1} (1-r)^{n-1-k} r^k (k - rn)/k, & 0 < k < n, \\ r^n, & k = n, \end{cases} \tag{9.12}$$

where b_{nk} is real and > 0 for $k > rn$. If we let $\alpha \equiv \alpha(n)$ be the greatest integer $\le rn$ and $| s_n | < M$, then again (8.13), (8.14), and (8.15) hold. Hence we have

$$S_n = 2 \sum_{k=0}^{\alpha} (a_{n-1,k} - a_{nk}) = 2\binom{n-1}{\alpha} (1-r)^{n-1-\alpha} r^{\alpha+1}, \tag{9.13}$$

and so

$$S_n \le \begin{cases} \dfrac{M(n-1)^{n-\frac{1}{2}}}{\left(n - 1 - \dfrac{1-r}{r}\right)^n}, & 0 < r \le \frac{1}{2}, \\[2ex] \dfrac{M(n-1)^{n-\frac{1}{2}}}{\left(n - 1 - \dfrac{r}{1-r}\right)^n}, & \frac{1}{2} \le r < 1. \end{cases} \tag{9.14}$$

Hence $S_n \to 0$ as $n \to \infty$, and we have

THEOREM 9.2. *The set of limit points of the Euler transform of a bounded complex sequence is connected for* $0 \le r < 1$.

[7] See, for example, W. A. Hurwitz, Bulletin of the American Mathematical Society, vol. 28(1922), p. 22.

Let us next consider the inverse Euler transformation for $r = a + ib$, denoted by

$$E_r^{-1} = E_{1/r}:\ s_n = \sum_{k=0}^{n} \binom{n}{k} \left(1 - \frac{1}{r}\right)^{n-k} \left(\frac{1}{r}\right)^{k} \sigma_k, \tag{9.21}$$

and the sequence

$$\sigma_n = \theta^n, \qquad \theta = e^{2\pi i/m} \qquad (n = 0, 1, 2, \cdots), \tag{9.22}$$

where m is an integer $\neq 0, \pm 1$. Substituting in (9.21) we have

$$s_n = \left(1 - \frac{1}{r} + \frac{\theta}{r}\right)^n, \tag{9.23}$$

which is bounded if and only if

$$\left| 1 - \frac{1}{r} + \frac{\theta}{r} \right| \leqq 1, \tag{9.24}$$

or if

$$a - b \cot \frac{\pi}{m} \geqq 1. \tag{9.25}$$

For each fixed r except r real and < 1, there exists an integer $m \neq 0, \pm 1$ such that (9.25) holds and $\{s_n\}$ is bounded. The set of limit points of the transform (9.22) contains exactly $| m |$ points, and so is not connected. Hence we have

THEOREM 9.3. *If r is any complex number except a real number < 1, then the set of limit points of the Euler transform (of order r) of a bounded sequence need not be connected.*

THE CITADEL.

FUNCTIONS OF BOUNDED VARIATION AND NON-ABSOLUTELY CONVERGENT INTEGRALS IN TWO OR MORE DIMENSIONS

BY R. L. JEFFERY

1. **Introduction.** This paper originated in an attempt to extend to two or more dimensions a property which completely characterizes a function of bounded variation in one dimension. It has been shown elsewhere[1] that a necessary and sufficient condition that a function $F(x)$ be of bounded variation is that there exist a summable function $f(x)$ and a sequence of summable functions $s_n(x)$ tending to $f(x)$ with $\int_e s_n\,dx$ bounded in n and e for which

$$F(x) = \lim_{n\to\infty} \int_a^x s_n\,dx. \tag{1}$$

We carry this idea over to a function $F(x, y)$ of the two real variables x and y by saying that $F(x, y)$ is in class V_1 on the rectangle $R = (0, 0; a, b)$ if there exist a single-valued function $f(x, y)$ and a sequence of summable functions $s_n(x, y)$ tending to $f(x, y)$ such that $\int_e s_n\,dx\,dy$ is bounded in n and e, and for which

$$F(x, y) = \lim_{n\to\infty} \int_0^y \int_0^x s_n\,dx\,dy. \tag{2}$$

The corresponding statement for functions of more than two variables is obvious. Since the function $f(x, y)$ is the limit of a sequence of measurable functions, it is measurable, and it will be shown that $f(x, y)$ is summable. Hence (2) is a direct extension of (1). It will also be shown that a function in class V_1 on R has what have come to be called bounded variation properties: It can be represented as the difference of two non-decreasing[2] functions; consequently the set of discontinuities of $F(x, y)$ has zero measure, and the surface $z = F(x, y)$ has a tangent plane almost everywhere; if ω is a rectangular interval with sides parallel to the coördinate axes, then the function of intervals $F(\omega)$ associated with $F(x, y)$ by the relation $F(\omega) = \lim \int_\omega s_n\,dx\,dy$ is additive,[3] of bounded variation, and

Received December 10, 1938; in revised form, October 11, 1939.

[1] Jeffery, *Functions defined by a sequence of integrals and the inversion of approximate derived numbers*, Transactions of the American Mathematical Society, vol. 41(1937), p. 175, Theorem IV.

[2] Non-decreasing in the sense of Hobson. If $x_2 \geq x_1$, $y_2 \geq y_1$, then $F(x_2, y_2) \geq F(x_1, y_1)$.

[3] A function of intervals $F(\omega)$ is additive on R if $F(\omega_1 + \omega_2) = F(\omega_1) + F(\omega_2)$ whenever ω_1, ω_2, and $\omega_1 + \omega_2$ are intervals on R. A function of sets $F(e)$ is additive if for every pair of disjunct sets e_1, e_2, $F(e_1 + e_2) = F(e_1) + F(e_2)$. $F(e)$ is completely additive if for every infinite sequence e_1, e_2, $\cdots$ of disjunct sets $F(\sum e_i) = \sum F(e_i)$.

$F'(x, y)$ the derivative[4] of $F(\omega)$ at the point (x, y) exists almost everywhere on R and is summable[5] over R. Since (2) is a direct extension of (1), and since $F(x, y)$ has the bounded variation properties just listed, it seems appropriate to call functions $F(x, y)$ in class V_1 on R functions of bounded variation in the two variables x and y. There are already in the literature several definitions for functions of bounded variation in two or more variables. An investigation of these definitions and the relations among them has been carried out by J. A. Clarkson and C. R. Adams.[6] Since functions in class V_1 can be written as the difference of two monotone functions, they are in class A (Arzela). It can also be shown that such functions are in class H (Hardy-Krause). Küstermann[7] has constructed an example which is in A but not in H. Hence $A > V_1$. For the present we must leave the question open as to whether or not $H = V_1$.

At this stage modifications in the definition of functions in class V_1 began to suggest themselves, e.g., the removal of the restriction that $\int_e s_n \, dx \, dy$ be bounded in n and e. Would this allow the introduction of associated functions $f(x, y)$ which were not summable? A consideration of these and related questions led straight to a theory of non-absolutely convergent integrals in two or more dimensions. It was with a view to the development of this theory that the following notation and definitions were laid down.

A function $F(x, y)$ in class V_1 on R is *completely additive over a measurable set* $E \subset R$ if $F(e) = \lim \int_e s_n \, dx \, dy$ exists and is completely additive, where e is the class of measurable sets $e \subseteq E$. It will be shown that a necessary and sufficient condition that $F(e) = \int_e f \, dx \, dy$ for every measurable set $e \subseteq E$ is that $F(x, y)$ is completely additive over E.

Let the function $f(x, y)$ be single valued on R, and let $s_n(x, y)$ be a sequence of summable functions tending to f. The function $F(x, y)$ is *in class* V_2 *on* R if

$$F(x, y) = \lim_{n \to \infty} \int_0^y \int_0^x s_n \, dx \, dy.$$

The function $f(x, y)$ is measurable, since it is the limit of a sequence of measurable functions. We also call attention to the fact that the function f need not be summable, nor is it necessary that $\int_e s_n \, dx \, dy$ be bounded in n and e.

[4] $F'(x, y)$ is the derivative of $F(\omega)$ at the point (x, y) if it is the limit as $m\omega$ tends to zero of $F(\omega)/(m\omega)$, where λ the ratio of the shorter side of ω to the longer side is bounded from zero. If there are no restrictions on λ, this limit is the strong derivative F'_s of $F(\omega)$.

[5] Saks, *Theory of the Integral*, Warsaw, 1937, pp. 115, 119.

[6] Transactions of the American Mathematical Society, vol. 35(1933), pp. 824–854; vol. 36(1934), pp. 711–730. A statement of the various definitions is given in the first of these papers.

[7] *Funktionen von beschränkter Schwankung in zwei reellen Veränderlichen*, Mathematische Annalen, vol. 77(1916), pp. 474–481.

A function $F(x, y)$ in class V_2 on R is *in class* V_1 *on a measurable set* $E \subset R$ if for every measurable set $e \subset E$ the limit of $\int_e s_n \, dx \, dy$ exists, and $\int_e s_n \, dx \, dy$ is bounded in n and e. It will be shown that if $F(x, y)$ is in class V_2 on R, and in class V_1 over $E \subset R$, then $f(x, y)$ is summable over E.

Let E be any closed set on R, and let $F(x, y)$ be a function in class V_2 on R. The function $F(x, y)$ is then *in class* V_1 *at a point* (x, y) *of* E, if for every sufficiently small interval ω containing (x, y) either as an interior point or boundary point the function $F(x, y)$ is in class V_1 over ωE. Let $\mathcal{C}(R, E)$ be a non-overlapping set of rectangles $c_1, c_2, \cdots$ (to be explicitly defined later) which covers CE, the complement of E. The function $F(x, y)$ in class V_2 on R is *completely additive at a point* (x, y) *of the closed set* E if, for every sufficiently small interval ω containing (x, y) either as an interior point or a boundary point, $F(e) = \lim \int_e s_n \, dx \, dy$ exists for every measurable set $e \subsetneq \omega E$, $F(e)$ is completely additive over ωE, and $F(\sum u_i) = \sum F(u_i)$, where u_i are the intervals and parts of intervals of the set $\mathcal{C}(R, E)$ which are on ω and $F(\sum u_i) = \lim \int_{\sum u_i} s_n \, dx \, dy$. $F(x, y)$ is *absolutely additive at* (x, y) if it is completely additive at (x, y) and $\sum w_i$ converges, where w_i is the least upper bound of $|F(\omega)|$ for ω on u_i.

Let $F(\omega)$ be a function of intervals on R, and (x, y) a point of R. If in any way a sequence of intervals ω_i can be specified which are such that $\omega_i \supset (x, y)$, the greatest diameter of ω_i tends to zero, and $F(\omega_i)/(m\omega_i)$ tends to a limit, then this limit is a *specialized derivative*, $F'_{sp}(x, y)$, of $F(\omega)$ at the point (x, y).

In terms of these definitions the following theorems will be proved:

THEOREM VIII. *If the function $F(x, y)$ is in class V_2 on R and is such that if E is any closed set on R then the points of E at which $F(x, y)$ is not of class V_1 and completely additive are non-dense on E, then $F'_{sp}(x, y)$ exists almost everywhere on R and is equal to $f(x, y)$.*

THEOREM IX. *If the function $F(x, y)$ is in class V_2 on R and is such that if E is any closed set on R then the points of E at which $F(x, y)$ is not of class V_1 and absolutely additive are non-dense on E, then $F'(x, y)$ exists almost everywhere on R and is equal to $f(x, y)$.*

If the continuous function $F(x, y)$ is in class V_2 on R, and satisfies the hypotheses of Theorem VIII, or of Theorem IX, with respect to a given function $f(x, y)$, then the function $F(x, y)$ is the *non-absolutely convergent integral* of $f(x, y)$. There are two motives back of this definition: The first is that if R is interpreted as a linear interval and $\mathcal{C}(R, E)$ is the set of open intervals complementary to the closed set E on R, then the function $F(x)$ corresponding to the function $F(x, y)$ in Theorem VIII is the Denjoy-Khintchine-Young integral of the function $f(x)$ which corresponds to $f(x, y)$, and F'_{sp} becomes the approximate derivative of $F(x)$. Similarly, in Theorem IX the function $F(x)$ becomes the

special Denjoy integral of $f(x)$, and $F'(x, y)$ becomes the derivative of $F(x)$. These statements we have proved elsewhere.[8] The second motive is that if $F(x, y)$ and $\varphi(x, y)$ are two functions which are continuous and satisfy the hypotheses of Theorem VIII, or of Theorem IX, with respect to a given function $f(x, y)$, then $F(x, y) = \varphi(x, y)$. It is easy to show this if we assume, what will be proved later, that if $F(x, y)$ is in class V_1 and completely additive over E, then $F(E) = \int_E f\,dx\,dy$. Let E_1 be the points of R at which at least one of the functions $F(x, y)$ and $\varphi(x, y)$ is not of class V_1 and completely additive. Then E_1 is non-dense on R, and if ω is an interval on CE_1, then $F(\omega) = \int_\omega f\,dx\,dy = \varphi(\omega)$. Let E_2 be the points of E_1 at which at least one of the functions $F(x, y)$ and $\varphi(x, y)$ is not of class V_1 and completely additive. Then E_2 is non-dense on E_1. Let ω be an interval on CE_2, and e the part of E_1 on ω. Then F and φ are of class V_1 and completely additive over e, and consequently $F(e) = \int_e f\,dx\,dy = \varphi(e)$. Furthermore, since $F(e)$ and $\varphi(e)$ are additive functions of sets,

$$F(\omega) = F(e) + F(\sum u_i), \qquad \varphi(\omega) = \varphi(e) + \varphi(\sum u_i),$$

where u_i are the intervals and parts of intervals of $\mathcal{C}(R, E_1)$ on ω. Also, since F and φ are completely additive at each point of E_1 on CE_2, it follows that

$$F(\sum u_i) = \sum F(u_i), \qquad \varphi(\sum u_i) = \sum \varphi(u_i).$$

In the definition of $\mathcal{C}(R, E)$ the intervals c_i, which are to have sides parallel to the coördinate axes, are not precluded from having points of E as boundary points. Hence an interval u_i may have points of E_1 on its boundary, and it is not known that $F(u_i) = \varphi(u_i)$. But if ω' is an interval interior to u_i, then $F(\omega') = \varphi(\omega')$, and on account of the continuity of F and φ it follows that as $\omega' \to u_i$, $F(\omega') \to F(u_i)$ and $\varphi(\omega') \to \varphi(u_i)$. It then follows that $F(u_i) = \varphi(u_i)$, and this, with the foregoing, shows that $F(\omega) = \varphi(\omega)$, where ω is any interval on CE_2. This argument can be continued. If a set E_α is not reached in a finite number of steps which is such that $E_{\alpha+1}$ is empty, let E_λ be the set common to all the sets $E_1, E_2, \cdots$. Then E_λ is closed, and every interval ω on CE_λ is on CE_n for some sufficiently great value of n. Consequently in the process of obtaining E_λ it has been shown that $F(\omega) = \varphi(\omega)$ for every interval ω on CE_λ. Let $E_{\lambda+1}$ be the points of E_λ at which at least one of the functions $F(x, y)$ and $\varphi(x, y)$ is not of class V_1 and completely additive. Then $E_{\lambda+1}$ is non-dense on E_λ and the foregoing method can be used to show that $F(\omega) = \varphi(\omega)$ for every interval ω on $CE_{\lambda+1}$. Since each of the sets $E_1, E_2, \cdots, E_\lambda, E_{\lambda+1}, \cdots$ is a proper subset of all the sets that precede it, there exists a set E_α, where α is a transfinite number of the first or second class, which is such that $E_{\alpha+1}$ is empty.[9]

[8] Bulletin of the American Mathematical Society, vol. 44(1938), pp. 840–845.

[9] Hobson, *Theory of Functions of a Real Variable*, vol. I, 3d edition, §82.

The set $CE_{\alpha+1}$ is the rectangle R, and at this stage it is possible to show that $F(\omega) = \varphi(\omega)$ for every interval ω on R.

We have thus given descriptive definitions of non-absolutely convergent integrals in two dimensions, and the corresponding statements for more than two dimensions are obvious. It is possible to give constructive definitions based on closed sets E and coverings $\mathcal{C}(R, E)$ of CE. We denote such an integral over an interval ω by $I(\omega, f)$, or simply $I(\omega)$, and state the definition as follows:

If f is summable over ω, then $I(\omega) = \int_\omega f\,dx\,dy$. If $I(\omega')$ exists for every interval ω' interior to an interval ω, then the limit as $\omega' \to \omega$ of $I(\omega')$ exists, and this limit is $I(\omega)$. Let E be any closed set on R which is such that $I(\omega)$ exists for every interval ω on CE. Let ω be an interval containing points of E for which f is summable over ωE, and for every interval ω' interior to or coincident with ω, let $\sum w_i'$ converge, where w_i' is the least upper bound of $|F(\omega)|$ for ω on u_i', the intervals and parts of intervals of $\mathcal{C}(R, E)$ on ω'. Then $I(\sum u_i') = \sum I(u_i')$, and $I(\omega') = \int_{\omega' E} f\,dx\,dy + \sum I(u_i')$. If f is such that this interval ω exists for every closed set E on R, the usual methods of transfinite induction apply to establish the existence of $I(\omega)$ for every interval ω on R.

That the function $I(0, 0; x, y)$ is in class V_2 on R can be shown by the methods used in the proof of Theorem XI, which theorem we now state:

THEOREM XI. *If the additive function of intervals $F(\omega)$ is continuous on R and is such that $F_s'(x, y)$ exists and is finite at each point of R, then $F(0, 0; x, y)$ is in class V_2 on R.*

With $I(0, 0; x, y)$ shown to be in class V_2 on R, it is evident from the construction of $I(\omega)$ that $I(0, 0; x, y)$ satisfies the conditions of Theorem IX. If in the definition of $I(\omega)$ we replace the convergence of $\sum w_i'$ by the convergence of $\sum |I(u_i')|$, we arrive at a constructive definition of $I(\omega)$ corresponding to Theorem VIII. We have not been able to show that for this definition $I(0, 0; x, y)$ is in class V_2, nor have we been able to replace $F_s'(x, y)$ by $F'(x, y)$ in Theorem XI.

In §4 we consider the problem of inverting derivatives of functions of intervals. In particular we prove

THEOREM X.[10] *If $F(\omega)$ is a continuous additive function of intervals defined on R, and F' is finite at each point of R, then it is possible to determine $F(\omega)$ in at most a denumerable set of operations on F'.*

[10] Results similar to this have been obtained by others, the first of whom was Looman, Fundamenta Mathematicae, vol. 4(1932), pp. 246-285. He defines

$$F'(x,y) = \lim_{h\to 0, k\to 0} \frac{F(x+h, y+k) - F(x+h, y) - F(x, y+k) + F(x, y)}{hk},$$

and shows that when F' is finite everywhere $F(x, y)$ can be determined in at most a denumerable set of operations. Later Krzyzansky and Kempisti, Comptes Rendus, Paris, vol. 198(1934), pp. 2058-2062, gave descriptive definitions of non-absolutely convergent integrals

We now introduce some further notation and some results which are essential in the proofs of the foregoing theorems.

Let $D = D_1, D_2, \cdots$ be a sequence of nets on R for which D_{n+1} is obtained by subdividing D_n, and for which the greatest diameter of any mesh of D_n tends to zero as n increases.

LEMMA I. *Let E be any closed set on R. If for each point (x, y) of E there exists n_{xy} such that if $n > n_{xy}$ any mesh of D_n which contains (x, y) has a property* P, *then there exists a finite set $K = k_1, k_2, \cdots$ of meshes of D which contain all of E either as interior points or boundary points, each of which has property* P, *and for which $mK - mE < \epsilon$.*

This lemma is of interest apart from its applications in the present paper on account of its close analogy to Vitali's covering theorem. In the present case there is no restriction on the ratio of the sides of the meshes of D. In Vitali's theorem, where the family of intervals covering E are not necessarily meshes of a sequence of nets, it is essential that λ, the ratio of the shorter side to the longer side of the meshes, be bounded from zero.[11]

If the sequence of nets D is further restricted to be such that λ is bounded from zero, it is possible to prove

LEMMA V. *Let $F(\omega)$ be a continuous additive function of intervals on R, and F' be finite at each point of R. Let E be the points of non-summability of F' over R. If F' is summable over E, there exists a sequence $c_1, c_2, \cdots$ of non-overlapping meshes of D which cover CE, and for which*

$$F(R) = \sum F(c_i) + \int_E F'\,dx\,dy.$$

Many of the foregoing results involve a consideration of a covering for CE, the complement of the closed set E on R. If R is interpreted as a linear interval there is a unique covering for CE, viz., the open intervals complementary to E. There is nothing comparable to this in the way of a unique covering in two or more dimensions. We complete the necessary notation for this paper with a description of a covering for CE which is completely determined when the two-dimensional interval R and the closed set E on R are given. In some situations a

based on the idea of generalized absolute continuity, and showed respectively that Looman's derivative and Looman's derivative restricted by $\alpha \leqq h/k \leqq 1/\alpha$ are integrable to $F(x, y)$ by their respective definitions. In the present paper there is an entirely different approach to the problem of inverting derivatives.

At this point we mention an interesting paper by Ridder, C. R. Soc. Sci., Varsovie, Class III, vol. 28(1935), pp. 5-16. Some of his results are comparable to results in the last section of the present paper, where strong derivatives are considered. Here also the methods are very different.

[11] Carathéodory, *Vorlesungen über reelle Funktionen*, Leipzig and Berlin, 1927, pp. 689-692.

simpler covering would serve; but the one described seems to have the minimum of essential properties for the work in the last section of this paper which deals with strong derivatives.

Operation $A(R, E)$. Let R be a rectangle with sides parallel to the coördinate axes, and let E be a closed set on R. Let e_1 be the left-hand end or the upper end of R according as the long sides of R are horizontal or vertical. Start at e_1, and by lines perpendicular to the long sides of R divide R into squares, or into squares and one rectangle with $\frac{1}{2} < \lambda < 1$, where this rectangle is in the end of R remote from e_1. Let $q_1, q_2, \cdots$ be the intervals of the finite set into which R has been divided which have points of E as interior points. If two or more of the remaining intervals of the set can be combined to form a single interval which has no points of E on its interior, let this be done, to obtain a set $h_1, h_2, \cdots$ of intervals on R which have no points of E as interior points. If R is a square, or a rectangle with $\frac{1}{2} < \lambda < 1$, no subdivisions are made. Nevertheless one still considers that the operation $A(R, E)$ has been performed.

For the set q_i we have $\frac{1}{2} < \lambda \leqq 1$. If h' is an interval of the set $h_1, h_2, \cdots$ which has no points of E as boundary points, then h' is adjacent to an interval q' of the set $q_1, q_2, \cdots$ and $mq' < 2mh'$.

Operation $B(R, E)$. Let the origin of coördinates be at the center of R. The coördinate axes divide R into four rectangles r_1, r_2, r_3, r_4, where the subscripts correspond to the quadrants. Let r'' be the first of these which has no points of E on its interior. Enlarge r'' to get r' by moving the perpendicular side of r'' parallel to itself until it encounters a point of E or a side of R. Enlarge r' to get r by moving the horizontal side of r' (not including end points) parallel to itself until it encounters a point of E or a side of R. There are now two cases to consider: (i) no corner of r interior to R; (ii) a corner K of r interior to R. In case (i) R is divided into rectangles r and $\bar{r}$. In case (ii) produce the perpendicular side of r through K until it intersects a side of R in a second point. Then R is divided into three rectangles r_1', r_2', r_3', where $r_1' = r$ and the ordering is counterclockwise about K. In both cases the rectangle r has at least one point of E on its boundary. If each of the rectangles r_1, r_2, r_3, r_4 has points of E on its interior, the operation $B(R, E)$ consists of dividing R into these four rectangles.

Given R and E the various intervals arising in the operations $A(R, E)$ and $B(R, E)$ are uniquely defined. We now proceed to define a covering for CE. Starting with R and E, perform the operation $A(R, E)$. Let Q_1 and H_1 be the sets $q_1^1, q_2^1, \cdots$ and $h_1^1, h_2^1, \cdots$ arising from this first application of $A(R, E)$. On each interval q of the set Q_1 perform the operation $B(q, E)$ and then perform the operations $A(\bar{r}, E)$, $A(r_2', E)$, $A(r_3', E)$ on all of the rectangles $\bar{r}, r_2', r_3'$ arising from any operation $B(q, E)$. Let Q_2 be the intervals $q_1^2, q_2^2, \cdots$ and H_2 be the intervals r arising from the operations $B(q, E)$, together with the set $h_1^2, h_2^2, \cdots$ which arise from any of the operations $A(\bar{r}, E)$, $A(r_2', E)$, $A(r_3', E)$. Continuing this process, we denote by $\mathcal{C}(R, E) = c_1, c_2, \cdots$ the intervals of the

sets $H_1, H_2, \cdots$. The set $\mathcal{C}(R, E)$ is uniquely defined in terms of R and E, and covers CE. For the set $\mathcal{C}(R, E)$ and the sets $Q_1, Q_2, \cdots$ the following properties hold:

A. Every interval c of the set $\mathcal{C}(R, E) = c_1, c_2, \cdots$ has a point of E on its boundary, or is adjacent to an interval q of the set $Q_1, Q_2, \cdots$ which has points of E on its interior and is such that $mq < 2mc$ and $q + c$ forms a rectangle.

B. Every interval q of any of the sets $Q_1, Q_2, \cdots$ has $\frac{1}{2} < \lambda < 1$. If E' is the part of E which is on some interval of the set Q_n for every n, then E' is closed and $m(E - E') = 0$. Every point of $E - E'$ is on a side of some interval of the set $\mathcal{C}(R, E)$.

2. **Functions of class** V_1. We recall that a function $F(x, y)$ is in class V_1 on R if $F(x, y) = \lim \int_0^y \int_0^x s_n \,dx\,dy$, where s_n is a sequence of summable functions tending to a measurable function $f(x, y)$, and $\int_e s_n \,dx\,dy$ is bounded in n and e. Further properties of functions in class V_1 are embodied in the following theorems.

THEOREM I. *If the function $F(x, y)$ is in class V_1 on R, then the associated measurable function $f(x, y)$ is summable.*

This follows from Fatou's Lemma.[12]

THEOREM II. *If the function $F(x, y)$ is in class V_1 on R, it is in class H on R.*

From the definition of $F(x, y)$ it follows that $F(x, 0) \equiv F(0, y) \equiv 0$. Let $\Delta = (x_i, x_{i+1}; y_j, y_{j+1})$, $x_{i+1} > x_i$, $y_{j+1} > y_j$, and

$$\Delta F = F(x_{i+1}, y_{j+1}) - F(x_{i+1}, y_j) - F(x_i, y_{j+1}) + F(x_i, y_j).$$

Then $\Delta F = \lim \int_\Delta s_n \,dx\,dy$, and it follows from the fact that $\int_e s_n \,dx\,dy$ is bounded in n and e that $\sum |\Delta F|$ is bounded. Consequently F is in H.

THEOREM III. *If the function $F(x, y)$ is in class V_1 on R, it can be written as the difference of two monotone functions on R.*

This follows from the fact that F is in H.[13]

The fact that if F is in V_1 it is in H permits the statement of the following two theorems.

THEOREM IV.[14] *A function $F(x, y)$ in class V_1 is totally differentiable almost everywhere.*

[12] Saks, loc. cit., p. 29.

[13] Adams and Clarkson, loc. cit., vol. 36(1934), p. 718, Theorem 6.

[14] Burkill and Haslam-Jones, *Notes on the differentiability of functions of two variables*, Journal of the London Mathematical Society, vol. 7(1932), pp. 297–305.

THEOREM V.[15] *A function $F(x, y)$ in class V_1 is continuous almost everywhere.*

The function $F(\omega)$ associated with $F(x, y)$ by the relation $F(\omega) = \lim \int_\omega s_n \, dx \, dy$ is of bounded variation, since otherwise $\int_e s_n \, dx \, dy$ would not be bounded in n and e. As we have pointed out in the introduction, $F'(x, y)$ exists and is summable. We next prove

THEOREM VI. *If $F(x, y)$ is in class V_1 and is completely additive on R, then $F(R) = \int_R F' \, dx \, dy$.*

Since F' is summable, it follows that at almost all points of R

$$(1) \qquad \frac{1}{m\omega} \int_\omega F' \, dx \, dy \to F'$$

as $m\omega \to 0$, provided the intervals ω are such that λ is bounded from zero. Hence for a point (x, y) for which (1) holds there exists a sequence of intervals $\omega_1, \omega_2, \cdots$ with $m\omega_i \to 0$ such that

$$(2) \qquad \left| \frac{F(\omega_i)}{m\omega_i} - \frac{1}{m\omega_i} \int_{\omega_i} F' \, dx \, dy \right| < \epsilon.$$

Since, for the intervals ω_i, λ is bounded from zero, it is possible to use Vitali's covering theorem to select a finite non-overlapping set $K_1 = \omega_1, \omega_2, \cdots$ of these intervals such that $mK_1 > \frac{1}{2}mR$. This argument can be repeated to get a similar set of intervals K_2 with $mK_2 > \frac{1}{2}mCK_1$, where K_2 is a set of non-overlapping intervals on CK_1. Getting a similar set K_3 on $C(K_1 + K_2)$, and continuing, we obtain the infinite sequence of sets of non-overlapping intervals $K = K_1, K_2, \cdots$. The intervals of these sets can be arranged in single order $k_1, k_2, \cdots$ and for each k_i (2) is satisfied. Hence

$$\left| \sum F(k_i) - \sum \int_{k_i} F' \, dx \, dy \right| < \epsilon mR.$$

We have $\sum mk_i = mK = mR$. Then, since $F(\omega)$ is completely additive and F' is summable, $\sum F(k_i) = F(\sum_i k_i) = F(K)$, and $\left| F(K) - \int_R F' \, dx \, dy \right| < \epsilon mR$. But $F(R) = F(K) + F(CK) = F(K)$, since $mCK = 0$ and $F(CK) = \lim \int_{CK} s_n \, dx \, dy$. Since ϵ is arbitrary, we conclude that $F(R) = \int_R F' \, dx \, dy$.[16]

[15] Adams and Clarkson, loc. cit., vol. 36(1934), p. 723.

[16] This result could also be obtained as follows: If $me = 0$, then $F(e) = \lim \int_e s_n \, dx \, dy = 0$. From this and the fact that $F(e)$ is completely additive it follows that $F(e)$ is absolutely continuous. (Saks, *Theory of the Integral*, Warsaw, 1937; definition of additivity, §5, p. 8, and definition of absolute continuity, §13, p. 30.) That $F(R) = \int_R F' \, dx \, dy$ follows from the absolute continuity of $F(e)$. (Saks, loc. cit., Theorem (7.8), p. 121.)

It is important to note that Saks' definition of additivity is what we are calling in this

THEOREM VII. *Let the function $F(x, y)$ be of class V_1 on R, and let E be any measurable set on R. A necessary and sufficient condition that $F(e) = \int_e f\,dx\,dy$ for every measurable set $e \subseteq E$ is that F be completely additive over E.*

The condition is necessary. By Theorem I f is summable on R. Let $e_1, e_2, \cdots$ be any sequence of disjunct measurable sets with $e_i \subset E$. If $F(e) = \int_e f\,dx\,dy$ for every $e \subseteq E$, it follows that

$$F(\sum e_i) = \int_{\sum e_i} f\,dx\,dy = \sum \int_{e_i} f\,dx\,dy = \sum F(e_i).$$

This shows that F is completely additive over E. To show that the condition is sufficient, let η be given, let $e(l, \eta)$ be the part of e for which $|s_n - f| < \eta$, $n \geqq l$. Let

$$\mathfrak{S}_1 = e(l_1, \eta), \qquad \mathfrak{S}_i = e(l_i, \eta) - e(l_{i-1}, \eta) \qquad (i > 1, l_i > l_{i-1}, l_i \to \infty).$$

Then $\sum m\mathfrak{S}_i = me$, and

$$F(\mathfrak{S}_i) = \lim_{n\to\infty} \int_{\mathfrak{S}_i} s_n\,dx\,dy = \lim_{n\to\infty} \int_{\mathfrak{S}_i} (f + t_n)\,dx\,dy = \int_{\mathfrak{S}_i} f\,dx\,dy + \eta_i,$$

where $|\eta_i| < \eta m\mathfrak{S}_i$. Since η can be taken arbitrarily small, $F(e) = F(\sum \mathfrak{S}_i) = \sum F(\mathfrak{S}_i)$, and, since f is summable, $\sum \int_{\mathfrak{S}_i} f\,dx\,dy = \int_e f\,dx\,dy$, we have $F(e) = \int_e f\,dx\,dy$.

3. **Functions of class V_2, and the proof of Theorems VIII and IX.** If the function $F(x, y)$ is of class V_2 on R, then $\int_e s_n\,dx\,dy$ is not necessarily bounded in n and e; and furthermore the associated measurable function $f(x, y)$ is not necessarily summable. We do have, however,

paper complete additivity. Furthermore, Saks' definition of absolute continuity is: (1) $F(e)$ is absolutely continuous if for every set e with $me = 0$ we have $F(e) = 0$. The usual definition is: (2) $F(e)$ is absolutely continuous if corresponding to $\epsilon > 0$ there is a $\delta > 0$ such that $|F(e)| < \epsilon$ when $me < \delta$. These definitions are equivalent if $F(e)$ is completely additive (Saks, loc. cit., Theorem (13.2), p. 31). They are not equivalent if $F(e)$ is additive without being completely additive, as a simple example shows: Let E_n be the interval $0 \leqq x \leqq n^{-1}$, and $e_n = E_n - E_{n+1}$. Let $s_n = n$ on E_n, and $s_n = 0$ elsewhere on $(0, 1)$. Let $F(e) = \lim \int_e s_n\,dx$. $F(e)$ is additive. For if e_1 and e_2 are any two disjunct measurable sets, then $F(e_1 + e_2) = F(e_1) + F(e_2)$. If e is any set with $me = 0$, then $F(e) = 0$, and thus $F(e)$ is absolutely continuous according to definition (1). Since $mE_n \to 0$ and $F(E_n) = 1$ for every n, it follows that $F(e)$ is not absolutely continuous according to definition (2). $F(\Sigma e_n) = 1$ and $\Sigma F(e_n) = 0$, and this shows that $F(e)$ is not completely additive.

THEOREM I$'$. *If the function $F(x, y)$ is of class V_2 on R, and of class V_1 over a measurable set $E \subset R$, then the associated measurable function $f(x, y)$ is summable on E.*

The proof is that of Theorem I with E replacing R.

We now prove Theorem VIII stated in the introduction. First let R be the closed set E of the theorem. Then the points E_1 of R at which $F(x, y)$ is not of class V_1 and completely additive are non-dense on R. If ω is an interval on CE_1, $F(x, y)$ is completely additive and of class V_1 on ω, and consequently it follows from Theorem VII that $F(\omega) = \int_\omega f\,dx\,dy$. It then follows that at almost all points of CE_1 $F' = f$. Next let E_2 be the points of E_1 at which $F(x, y)$ is not of class V_1 and completely additive. Let e be the points of E_1 on CE_2 which are on an infinite sequence $q_1, q_2, \cdots$ of intervals of the sets $Q_1, Q_2, \cdots$ arising in the construction of $\mathcal{C}(R, E_1)$. Since $F(x, y)$ is of class V_1 and completely additive at a point (x, y) of e, there exists q_i on CE_2 containing (x, y) such that f is summable over $E_1 q_i$, and $\sum_j F(c_j^i) = F(\sum_j c_j^i)$, where c_j^i are the intervals of $\mathcal{C}(R, E_1)$ on q_i (there are only whole intervals of $\mathcal{C}(R, E_1)$ on q_i). Then

$$\frac{F(q_i)}{mq_i} = \frac{\int_{E_1 q_i} f\,dx\,dy}{mq_i} + \frac{\sum_j F(c_j^i)}{mq_i}. \tag{1}$$

For almost all points of e the first term on the right tends to f as mq_i tends to zero. We show that for almost all of e the second term on the right tends to zero. Suppose there exists $e' \subset e$ with $\overline{m}e' > 0$, $\overline{m}e'$ the outer measure of e', for which

$$\overline{\lim_{mq_i \to 0}} \frac{\sum_j |F(c_j^i)|}{mq_i} > d > 0. \tag{2}$$

Since $F(x, y)$ is completely additive at each point of e', there exists an interval ω with $\overline{m}\omega e' > 0$ such that if c_k are the intervals of $\mathcal{C}(R, E_1)$ with points on ω, then $\sum F(c_k) = F(\sum c_k)$. Since $F(\sum c_k)$ is independent of the ordering of the intervals c_k, it follows that $\sum |F(c_k)|$ converges. Hence there exists $\delta > 0$ such that if c_k' are the intervals of c_k with $mc_k' < \delta$, then

$$\sum |F(c_k')| < d\overline{m}\omega e'. \tag{3}$$

It follows from (2) that each point of $\omega e'$ is on an infinite sequence q_i of the intervals of the sets $Q_1, Q_2, \cdots$ which arise in the construction of $\mathcal{C}(R, E_1)$, for which $mq_i < \delta$ for all i, mq_i tends to zero, and $\sum_j |F(c_j^i)| > dmq_i$. Since for each q_i we have $\frac{1}{2} < \lambda \leq 1$, Vitali's covering theorem can be used to select a non-overlapping set q_i' of these intervals q_i with $\sum mq_i' > \overline{m}\omega e' - \epsilon$. Then from (2)

$$\sum_i \sum_j |F(c_j^i)| > d \sum mq_i' > d(\overline{m}\omega e' - \epsilon).$$

Since $mq_i' < \delta$ for every i, it follows that $mc_j^i < \delta$ for all i and j. Hence from (3) $\sum_i \sum_j | F(c_j^i) | < d\bar{m}\omega e'$. Since ϵ is arbitrary this gives a contradiction, and the desired result is established. We can now conclude that $F(q_i)/(mq_i) \to f$ for almost all of e', and consequently for almost all of E_1 on CE_2. In other words, at almost all of CE_2 the special derivative F'_{sp} exists and is equal to f. This process can be continued, as in the transfinite argument above, to establish the existence of $F'_{sp} = f$ for almost all points of R.

At this point the question arises as to the possibility of proving the existence of F'. This involves replacing q_i in (1) by ω_i with λ bounded from zero, and c_j^i by u_j^i. This could be done in such a way that $\sum_j | F(u_j^i) |$ converges for each i, but the conditions of Theorem VIII do not insure that this convergence is uniform in i. In Theorem IX it is assumed that $\sum w_i$ converges. This is sufficient to permit the necessary modifications in the proof of Theorem VIII to establish the existence of F' almost everywhere for functions satisfying the conditions of Theorem IX.

4. **The inversion of $F'(x, y)$.** In this section we make constant use of a sequence of nets $D = D_1, D_2, \cdots$ on R where D_{n+1} is obtained by subdividing D_n, where the greatest diameter of the meshes of D_n tends to zero as n increases, and where λ the ratio of the shorter side to the longer side of the meshes of D is bounded from zero. For the proof of Lemma I, as was stated in the introduction, this restriction on λ is not necessary; but it is essential for some of the results based on Lemma I.

To proceed with the proof of Lemma I we first put E in a set of open regions O with $mO - mE < \epsilon$. Each point of E is on the interior of a mesh G of some D_n where the mesh G has property P and is on O, or is on the interior of a region G formed by a group of not more than four meshes of some D_n where each mesh of the group has property P and is on O. Thus each point of E is on at least one region of a set of open regions which are on O. The Heine-Borel theorem[17] can now be used to select a finite set $O_1, \cdots, O_p$ of these regions which are on O, and which contain all of E. Two or more of these regions may have points in common. They can, however, be reduced to a non-overlapping set in the following way: Let $n_1, n_2, \cdots, n_q$ be the values of n arranged in increasing order of magnitude which are such that there are meshes of D_n in the sets $O_1, \cdots, O_p$. Let $H_1, \cdots, H_q$ represent the meshes of $D_{n_1}, \cdots, D_{n_q}$ respectively which go to make up the regions $O_1, \cdots, O_p$. Since $n_1 < n_2 < \cdots < n_q$, every mesh of the sets $H_2, \cdots, H_q$ is either wholly interior to or wholly exterior to the meshes of the set H_1. Delete from $H_2, \cdots, H_q$ all meshes which are on H_1. We thus arrive at the set $H_1, H_2', \cdots, H_q'$. Now delete from $H_3', \cdots, H_q'$ all meshes which are on any interval of the set H_2'. This gives the set $H_1, H_2', H_3'', \cdots, H_q''$. Continuing this process, we arrive at the set

[17] Hobson, *Theory of Functions of a Real Variable*, 3d edition, vol. 1, p. 109.

$H_1, H_2', H_3'', \cdots, H_q^{(q-1)}$ of non-overlapping regions which contain all of E, which are such that each region consists of a group of not more than four meshes of D, and such that each mesh of any group has property P and is on O. If the separate meshes of these regions are denoted by $K = k_1, k_2, \cdots$, then the set K satisfies the requirements of the lemma.

LEMMA II. *Let D be a sequence of nets on R with λ bounded from zero. Let $F(\omega)$ be an additive function of intervals defined on R such that F' is finite at each point of a closed set E and summable over E. Corresponding to a given $\epsilon > 0$ there exists a finite set $K = k_1, k_2, \cdots$ of non-overlapping meshes of D which are such that each point of E is on K, either as an interior point or as a boundary point, and for which*

$$\sum | F(k_i) | < \int_E | F' | \, dx \, dy + \epsilon \qquad \textit{and} \qquad mK - mE < \epsilon.$$

Let $a_0, a_1, \cdots$ be a subdivision of $(0, \infty)$, e_i the part of E for which $a_{i-1} \leqq | F' | < a_i$, and α_i an open set containing e_i with $m\alpha_i - me_i < \epsilon_i$. A mesh k of D which contains a point (x, y) of e_i has property P if k is on α_i and $|F(k)|/(mk) < a_i$. Lemma I then applies to give a finite non-overlapping set $K = k_1, k_2, \cdots$ of intervals of D which contains E and each interval of the set K has property P. Hence

$$\sum | F(k_i) | < \sum a_i mk_i < \sum a_i m\alpha_i < \sum (me_i + \epsilon_i)a_i .$$

The lemma follows if first a_i and then ϵ_i are suitably chosen.

LEMMA III. *If D is a sequence of nets with λ bounded from zero, and F' is finite at each point of a closed set E and summable over E, there exists a finite non-overlapping set $K = k_1, k_2, \cdots$ of meshes of D for which*

$$\left| \sum F(k_i) - \int_E F' \, dx \, dy \right| < \epsilon, \qquad | mE - mKE | < \epsilon, \qquad \textit{and} \qquad | mK - mE | < \epsilon.$$

To prove this we let (a_{i-1}, a_i) be a subdivision of $(-\infty, \infty)$, e_i the part of E for which $a_{i-1} \leqq F' < a_i$, and α_i an open set containing e_i. Since λ is bounded from zero for the set of meshes of D, there exists about each point (x, y) of e_i an infinite sequence k_j^i of meshes of D such that k_j^i is on α_i, $mk_j^i \to 0$, and $a_{i-1} - \epsilon_i < F(k_j^i)/(mk_j^i) < a_i$. Vitali's covering theorem can then be used to select a finite non-overlapping set of these intervals with $| \sum_j mk_j^i - me_i | < \epsilon_i$, and $| me_i - \sum_j mk_j^i e_i | < \epsilon_i$. With adequate attention to detail, this process can be carried out for a finite number of the sets e_i in such a way that the sets k_j^p and k_j^q have no points except boundary points in common if $p \neq q$. We then have

$$\sum_i \sum_j mk_j^i (a_{i-1} - \epsilon_i) < \sum_i \sum_j F(k_j^i) < \sum_i \sum_j mk_j^i a_i .$$

If the quantities a_i and ϵ_i are suitably chosen, this inequality shows that

$$\left| \sum_i \sum_j F(k_j^i) - \int_E F' \, dx \, dy \right| < \epsilon,$$

and if $K = \sum_i \sum_j k_j^i$, then $| mE - mKE | < \epsilon$, and $| mK - mE | < \epsilon$. This establishes the lemma.

LEMMA IV. *If $F(\omega)$ is an additive function of intervals on R and F' is finite at each point of R and summable over R, then $F(R) = \int_R F' \, dx \, dy$.*

Let ϵ be given, and let δ be fixed so that if $me < \delta$ then $\int_e | F' | \, dx \, dy < \epsilon$. Let D be a sequence of nets on R with the greatest diameter of the meshes tending to zero and with λ bounded from zero. It follows from Lemma III that there exists a finite non-overlapping set $K = k_1, k_2, \cdots$ of meshes of D with $mR - mK < \delta$ and

$$\left| \sum F(k_i) - \int_R F' \, dx \, dy \right| < \epsilon .$$

Let G be the closed set complementary to the interior points of K, and let n' be the greatest value of n for which there are meshes of the net D_n of D in the set K. It now follows from Lemma II that there exists a finite non-overlapping set $g' = g_1', g_2', \cdots$ of meshes of D chosen from nets D_n with $n > n'$, where g' contains all of G either as interior points or boundary points, and

$$\sum | F(g_i') | < \int_G | F' | \, dx \, dy + \epsilon.$$

Since $n > n'$ every mesh of g' is wholly interior to K or wholly exterior to K, except possibly for boundary points. If the meshes of g' which are interior to K are deleted from g', there remains a set $g = g_1, g_2, \cdots$ with the same properties relatively to the closed set G as the set g'. Since $F(\omega)$ is additive, $F(R) = \sum F(k_i) + \sum F(g_i)$, and consequently

$$\left| F(R) - \int_R F' \, dx \, dy \right| < 2\epsilon.$$

Since ϵ is arbitrary, we have $F(R) = \int_R F' \, dx \, dy$,[18] and this is the lemma.

We complete this sequence of results with the proof of Lemma V stated in the introduction. Applying Lemma III we get the finite set of meshes $K =$

[18] It is to be noted that this fundamental result has been obtained without any assumption as to the absolute continuity, continuity, or complete additivity of $F(\omega)$. Compare with Theorem VI, where F' is not required to be finite. That the function of intervals $F(\omega)$ is absolutely continuous follows from Lemma II, and this could be used to lead to a proof of Lemma IV.

$k_1, \cdots, k_n$ of D with

$$\left| \sum F(k_i) - \int_E F' \, dx \, dy \right| < \epsilon_1,$$

and $mE - \sum mk_i E < \epsilon_1$. Let E' be the part of E which is exterior to and on the boundary of the set $K = k_1, \cdots, k_n$. Then E' is closed and $mE' < \epsilon_1$. It is now possible to apply Lemma II and get a finite set $l_1, l_2, \cdots$ of meshes of D which contain all of E' either as interior points or boundary points and for which

$$\sum | F(l_i) | < \int_{E'} | F' | \, dx \, dy + \epsilon_1,$$

with $| \sum ml_i - mE' | < \epsilon_1$. Furthermore, if n' is the largest value of n for which the net D_n of the sequence of nets D has meshes in the set $k_1, \cdots, k_n$, then every mesh of the set $l_1, l_2, \cdots$ can be chosen from nets D_n with $n > n'$. This insures that no mesh of the set $l_1, l_2, \cdots$ contains a mesh of the set $k_1, \cdots, k_n$. Drop from $l_1, l_2, \cdots$ all meshes which are contained by a mesh of the set $k_1, \cdots, k_n$. The remaining set will still satisfy

$$\sum | F(l_i) | < \int_{E'} | F' | \, dx \, dy + \epsilon_1 .$$

Let $k_1^1, \cdots, k_p^1$ be the combined sets $k_i + l_i$. Then $k_1^1, \cdots, k_p^1$ contains all of E either as interior points or boundary points and

$$\left| \sum F(k_i') - \int_E F' \, dx \, dy \right| < 2\epsilon_1,$$

provided the set $k_1, k_2, \cdots$ has been so chosen that $\int_{E'} | F' | \, dx \, dy < \epsilon_1$. Now let n' be the largest value of n for which D_n has meshes in the set $k_1^1, \cdots, k_p^1$. Take any net D_n with $n > n'$, and let $c_1^1, c_2^1, \cdots$ be the finite number of meshes of D_n which are not interior to k_i^1. Then $\sum c_i^1 + \sum k_i^1 = R$. A mesh c_i^1 has no points of E on its interior. Let ω' be an interval interior to c_i^1. Then $F(\omega') = \int_{\omega'} F' \, dx \, dy$, and $F(c_i^1) = \lim F(\omega')$ as $\omega' \to c_i^1$. Hence $F(c_i^1)$ can be determined from F', and we have

(1) $$F(R) = \sum F(c_i^1) + \sum F(k_i^1).$$

Consequently

$$F(R) = \sum F(c_i^1) + \int_E F' \, dx \, dy + \eta_1,$$

where $| \eta_1 | < 3\epsilon_1$. This process can be repeated on the set k_i^1 to get c_i^2 and k_i^2 both on k_i^1 with $\sum c_i^2 + \sum k_i^2 = \sum k_i^1$, and

$$\left| \sum F(k_i^2) - \int_E F' \, dx \, dy \right| < 3\epsilon_2 .$$

We then have

$$F(R) = \sum F(c_i^1) + \sum F(c_i^2) + \sum F(k_i^2), \tag{2}$$

and consequently

$$F(R) = \sum F(c_i^1) + \sum F(c_i^2) + \int_E F' \, dx \, dy + \eta_2,$$

where $|\eta_2| < 2\epsilon_2$. If the sequence $\epsilon_1, \epsilon_2, \cdots$ is such that $\epsilon_i \to 0$ and c_1, $c_2, \cdots$ is the sequence $c_1^1 \cdots c_{n_1}^1, c_1^2 \cdots c_{n_2}^2, \cdots$, then

$$F(R) = \sum F(c_i) + \int_E F' \, dx \, dy. \tag{3}$$

It should be emphasized here that the ordering of the sequence $c_1, c_2, \cdots$ cannot be changed.

We are now in a position to complete the proof of Theorem X stated in the introduction. In the first place, if E is any closed set on R, the points of E at which F' is not bounded over E are non-dense on E. This is obvious if E consists of isolated points and the limit points of isolated points. Let E be perfect, and let ω be an interval containing points of E on its interior. If the assertion is not true, there exists a point P of E interior to ω, and an interval ω_1 containing P on its interior with $|F(\omega_1)|/(m\omega_1) > M_1$. Similarly there exists ω_2 interior to ω_1 containing points of E and such that $|F(\omega_2)|/(m\omega_2) > M_2$. This process can be continued with $M_n \to \infty$, $\omega_{n-1} \supset \omega_n$, $m\omega_n \to 0$, and λ bounded from zero for the set $\omega_1, \omega_2, \cdots$. If (ξ, η) is the point common to these intervals, then (ξ, η) belongs to E, and at this point F' is infinite or does not exist. This contradicts the fact that F' is finite at each point of E, and the assertion is proved. It follows at once from this that the points of E at which F' is not summable over E are non-dense on E. Let E_1 be the points of non-summability of F' over R. Then E_1 is non-dense on R, and if ω is an interval on CE_1, then by Lemma IV $F(\omega) = \int_\omega F' \, dx \, dy$. Let E_2 be the points of non-summability of F' over E_1. Then E_2 is non-dense on E_1, and if ω is an interval on CE_2 containing points of E_1, it is possible to place a sequence of nets D on ω and determine $F(\omega)$ by the methods used in the proof of Lemma V. This process can be continued by finite and transfinite induction to show that for any interval ω on R, $F(\omega)$ can be determined in at most a denumerable set of operations, and this is Theorem X.

5. **The inversion of strong derivatives and the proof of Theorem XI.** As a preliminary to the proof of Theorem XI we give a method for inverting the strong derivative of a continuous additive function of intervals $F(\omega)$ which is the analogue of the Denjoy process for functions of a single variable. Let E be a closed set on R. Let $c_1, c_2, \cdots$ be the intervals of $\mathcal{C}(R, E)$, and w_i the least upper bound of $|F(\omega)|$ for ω on c_i. A point P of E is a point of convergence of $\sum w_i$ if there exists an interval ω about P for which $\sum w_i'$ con-

verges, where w_i' is the least upper bound of $|F(\omega')|$ for ω' on ωc_i. If no such interval exists, then P is a point of divergence of $\sum w_i$. We prove

Lemma VI. *Let $F(\omega)$ be a continuous additive function of intervals for which F_s' is finite at each point of R. If E is any closed set on R, then the points of E which are not points of summability of F_s' and points of convergence of $\sum w_i$ are non-dense on E.*

For the first part of the assertion we refer to the proof of Theorem X. The second part is obvious if E consists of isolated points and the limit points of isolated points. Let E be perfect, and let ω be an interval containing a point P of E on its interior. If $\sum w_i$ diverges at P, then $w_i/(mc_i)$ is unbounded at P. Hence we can find c_i interior to ω such that the distance between the boundary of c_i and the boundary of ω is not less than twice the greatest diameter of c_i, and such that if M_1 is an arbitrary positive number $w_i/(mc_i) > M_1$. This implies the existence of an interval r on c_i with $|F(r)|/(mc_i) > M_1$. From the manner of construction of $\mathcal{C}(R, E)$, the interval c_i either has a point of E on its boundary, or forms with some interval q_i of the sets $Q_1, Q_2, \cdots$ which enter into the construction of $\mathcal{C}(R, E)$ an interval $c_i + q_i$ which has a point of E on its interior. Furthermore, $mq_i < 2mc_i$, and from this and the above distance condition between c_i and ω, it follows that $c_i + q_i$ is interior to ω. If a point P of E is fixed on the boundary of c_i, or interior to q_i, then one possibility is a rectangle r' with one corner at P and with the rectangle r for which $F(r)/(mc_i) > M_1$ in the corner diagonally opposite to P, with r' on $c_i + q_i$, and consequently interior to ω. Let the origin of coördinates be at P, let $r' = (0, 0; x_2, y_2)$, and $r = (x_1, y_1; x_2, y_2)$. Then

$$(1)\qquad r = (0, 0; x_2, y_2) - (0, 0; x_2, y_1) - (0, 0; x_1, y_2) + (0, 0; x_1, y_1).$$

We now have $|F(r)|/(mc_i) > M_1$, r' on $c_i + q_i$, $mq_i < 2mc_i$, and consequently $|F(r)|/(mr') > \frac{1}{3}M_1$. Hence for at least one of the rectangles r'' on the right side of (1), $|F(r'')|/(mr') > \frac{1}{12}M_1$. Other possible situations can be handled in a similar way. Then, since $r' \supset r''$, $|F(r'')|/(mr'') > \frac{1}{12}M_1$. The interval r'' is interior to ω and has one corner at P. From this and the continuity of $F(\omega)$ it follows that there exists an interval ω_1 interior to ω, containing points of E on its interior, for which $|F(\omega_1)|/(m\omega_1) > \frac{1}{12}M_1$. The foregoing reasoning can be repeated on ω_1, and the process continued to get $|F(\omega_n)|/(m\omega_n) > \frac{1}{12}M_n$, with $M_n \to \infty$, $\omega_n \supset \omega_{n+1}$, and the greatest diameter of ω_n tending to zero. If then P' is the point contained in every interval of the set ω_n, P' belongs to E, and at P' either F_s' is infinite or does not exist. This is a contradiction, and the truth of the assertion follows.

Lemma VII. *If the interval ω contains points of E, if F_s' is summable over ωE and if $\sum |F(u_i)|$, where u_i are the intervals and parts of intervals of $\mathcal{C}(R, E)$ on ω, converges, then*

$$F(\omega) = \sum F(u_i) + \int_{\omega E} F'\,dx\,dy.$$

Let e' be the points of ωE which are on the sides of the intervals u_i but which are not limit points of the set u_i. Let $e = \omega E - e'$. Then e is closed, and $me = m\omega E$. Let $v_1^n, v_2^n, \cdots$ be the intervals and parts of intervals on ω of the set Q_n which enter at the n-th stage in the construction of $\mathcal{C}(R, E)$. Then each point of e is on an interval of the set $v_1^n, v_2^n, \cdots$ for all values of n. At the n-th stage in the construction of $\mathcal{C}(R, E)$ the finite set $u_1, \cdots, u_p$, $v_1^n, \cdots, v_q^n$ forms a net D_n on ω, and D_{n+1} is obtained by subdividing $v_1^n, \cdots, v_q^n$ and leaving $u_1, \cdots, u_p$ unchanged. The sequence of nets $D = D_1, D_2, \cdots$ thus obtained differs from the sequence used in Lemmas I-IV in that the greatest diameter of all the meshes of D_n does not tend to zero as n increases, and λ is not bounded from zero for the whole set of meshes. Nevertheless, the sequence of meshes $v_1^n, v_2^n, \cdots$ does satisfy both these conditions, and each point of the closed set e is on a mesh of this sequence for all values of n. It then follows that the methods of Lemmas I-IV are available to select a finite set $v_1, \cdots, v_m$ of the meshes of the sets $v_1^n, v_2^n, \cdots$ which contain all of e either as interior points or as boundary points, and for which $\left|\sum v_k - me\right| < \epsilon$ and $\left|\sum F(v_k) - \int_e F_s' \, dx\, dy\right| < \epsilon$. If now the construction of $\mathcal{C}(R, E)$ is modified to the extent of not subdividing an interval of any Q_n which is an interval of the set $v_1, \cdots, v_m$ or which overlaps ω to form an interval of this set $v_1, \cdots, v_m$, a stage is reached in a finite number of steps in the modified construction of $\mathcal{C}(R, E)$ at which

$$\omega = u_1 + \cdots + u_l + v_1 + \cdots + v_m,$$

where $u_1, \cdots, u_l$ are meshes and parts of meshes of $\mathcal{C}(R, E)$. We then have

$$F(\omega) = \sum_{i=1}^{l} F(u_i) + \sum_{i=1}^{m} F(v_i), \tag{1}$$

and hence

$$\left| F(\omega) - \sum_{i=1}^{l} F(u_i) - \int_e F_s' \, dx\, dy \right| < \epsilon.$$

The number ϵ is arbitrary, $\sum F(u_i)$ converges, and the intervals $v_1, \cdots, v_m$ can be chosen in such a way that the smallest value of n for which any interval of this set belongs to Q_n is arbitrarily great. As this value of n increases, the first term on the right side of (1) tends to include all the intervals of the set u_i. From these considerations we have

$$F(\omega) = \sum F(u_i) + \int_{\omega E} F_s' \, dx\, dy,$$

and the lemma is proved.

THEOREM X'. *Let F_s' be the strong derivative of a continuous additive function of intervals and finite at each point of R. Then $F(\omega)$ can be determined from F_s' by a process which is the analogue of the Denjoy process for inverting the finite derivative $F'(x)$ of a function $F(x)$ of a real variable x.*

By Lemma VI the points E_1 of non-summability of F'_s over R are non-dense on R. If then ω is an interval on CE_1, it follows from Lemma IV that $F(\omega) = \int_\omega F'_s\,dx\,dy$. If E_2 is the part of E_1 at which we do not have both F'_s summable over E_1 and $\sum w_i$ converging, where w_i is the least upper bound of $|F(\omega)|$ for ω on the interval c_i of the set $\mathcal{C}(R, E_1)$, then by Lemma VI E_2 is non-dense on E_1. If then ω is an interval on CE_2, it follows from Lemma VII that

$$F(\omega) = \sum F(u_i) + \int_{\omega E_1} F'_s\,dx\,dy,$$

where u_i are the intervals and parts of intervals of the covering $\mathcal{C}(R, E_1)$ which are on ω. The argument now proceeds as in the previous cases in which we have used transfinite induction to the stage where it is possible to determine $F(\omega)$ for every interval ω on R. A comparison with the methods developed by Denjoy[19] reveals that the two processes are the same.

We are now in a position to concern ourselves directly with the proof of Theorem XI. We first prove

LEMMA VIII. *Let $F(\omega)$ be a continuous additive function of intervals on R, and E a closed set on R. Relatively to these let the following conditions hold:* (1) *F'_s is finite at each point of E and summable over E.* (2) *$\sum w_i$ converges, where w_i is the least upper bound of $|F(\omega)|$ for ω on c_i, an interval of $\mathcal{C}(R, E)$.* (3) *The sequence $s_1, s_2, \cdots$ of summable functions is defined on CE in such a way that $s_n \to F'_s$, and if ω is an interval on CE then for ω' on ω $\int_{\omega'} s_n\,dx\,dy \to F(\omega')$ uniformly in ω'. If these conditions hold, there exists a sequence of summable functions on R with $s_n \to F'_s$ and such that for the intervals ω on R $\int_\omega s_n\,dx\,dy \to F(\omega)$ uniformly in ω.*

By Lemma VII

$$F(R) = \sum F(c_i) + \int_E F'_s\,dx\,dy.$$

Let $c_1^k, \cdots, c_k^k$ be intervals interior to $c_1, \cdots, c_k$, respectively, for which the following hold: (4) $c_i^k \to c_i$, $c_i^{k+1} \supset c_i^k$. (5) If ω is any interval on c_i, then $|F(\omega c_i) - F(\omega c_i^k)| < \epsilon_i^k$, where $\sum_i \epsilon_i^k = \epsilon_k$, $\epsilon_k < \epsilon_{k-1}$, and $\epsilon_k \to 0$. With k fixed the continuity of $F(\omega)$ permits this. Let $s_{nk} = s_n$ on c_i^k, $s_{nk} = F'_s$ on the part of E for which $-n < F'_s < n$, $s_{nk} = 0$ elsewhere on R. If k is fixed and n is sufficiently great, (3) gives

$$\left| F(\omega) - \int_\omega s_{nk}\,dx\,dy \right| < \epsilon_i^k \tag{6}$$

[19] See Hobson, *Theory of Functions of a Real Variable*, 3d edition, vol. 1, §§464, 465.

for ω on c_i^k. If ω is any interval on R, Lemma VII gives

$$F(\omega) = \sum F(u_i) + \int_{\omega E} F'_s \, dx \, dy, \tag{7}$$

where u_i are the intervals and parts of intervals of $\mathcal{C}(R, E) = c_1, c_2, \cdots$ on ω. If u_i^k is the part of u_i on c_i^k, then from (4) $u_i^k \to u_i$; from (6) $\left| F(u_i^k) - \int_{u_i^k} s_{nk} \, dx \, dy \right| < \epsilon_i^k$; from (5) $| F(u_i) - F(u_i^k) | < \epsilon_i^k$; from (2) $\sum_{k+1}^{\infty} | F(u_i) | = \eta_k$, where $\eta_k \to 0$ as $k \to \infty$; from the definition of s_{nk} on E,

$$\left| \int_{\omega E} s_{nk} \, dx \, dy - \int_{\omega E} F'_s \, dx \, dy \right| = \epsilon_n,$$

where $\epsilon_n \to 0$ as $n \to \infty$, uniformly in ω; these considerations, together with (7), show that

$$\left| F(\omega) - \int_{\omega} s_{nk} \, dx \, dy \right| < 2 \sum_i \epsilon_i^k + \eta_k + \epsilon_n$$

uniformly in ω, where $\sum_i \epsilon_i^k$, η_k, ϵ_n tend to zero as n and k tend to infinity. As things now stand, when n and k tend to infinity, $s_{nk} \to F'_s$ almost everywhere, the exceptional set being the boundaries of the rectangles of the set $\mathcal{C}(R, E) = c_1, c_2, \cdots$. On this exceptional set change the definition of s_{nk} so that $s_{nk} = F'_s$ for all values of n and k. If now a single sequence $s_{nk} = s_n$ is chosen where $k_n \to \infty$ as $n \to \infty$, this sequence s_n satisfies the requirements of the lemma.

Now let $E_1, E_2, \cdots, E_\lambda, E_{\lambda+1}, \cdots$ be the sets entering into the proof of Theorem X′. If $s_n = F'_s$ on the part of CE_1 for which $-n < F'_s < n$, and $s_n = 0$ elsewhere on R, then this sequence satisfies condition (3) of Lemma VIII relatively to the set E_1. Let $D = D_1, D_2, \cdots$ be a sequence of nets on R with the greatest diameter of the meshes of D_n tending to zero. Let G_k be the set $d_1^k, d_2^k, \cdots$ of meshes of D_k which do not have points of E_2 either as interior points or as boundary points. On each interval d_i^k of the set G_k all the conditions of Lemma VIII are satisfied, d_i^k replacing R, and $E_1 d_i^k$ replacing E. Then, since the intervals of G_k are finite in number, there exists a sequence s_{nk} on G_k with $s_{nk} \to F'_s$ and such that for the intervals ω on G_k $\int_\omega s_{nk} \, dx \, dy \to F(\omega)$ as $n \to \infty$, uniformly in ω. If now $s_n = s_{n1}$ on G_1, $s_n = s_{n2}$ on $G_2 - G_1$, $s_n = s_{n3}$ on $G_3 - G_2, \cdots$, and $s_n = 0$ on E_2, then this sequence satisfies condition (3) of Lemma VIII relatively to CE_2. This process can be continued. Let E_λ be the set common to all the sets $E_1, E_2, \cdots$. On CE_k there exists a sequence $s_{nk} \to F'_s$ such that, for ω on CE_k and for ω' the intervals on ω, $\int_{\omega'} s_{nk} \, dx \, dy \to F(\omega')$ uniformly in ω' as $n \to \infty$. Let $\mathfrak{S}_1 = CE_1$, $\mathfrak{S}_2 =$

$CE_2 - CE_1$, $\mathfrak{S}_3 = CE_3 - CE_2$, $\cdots$. Let $s_n = s_{n1}$ on $\mathfrak{S}_1$, $s_n = s_{n2}$ on $\mathfrak{S}_2$, $\cdots$. Then s_n is defined on CE_λ. We have $CE_k = \mathfrak{S}_1 + \mathfrak{S}_2 + \cdots + \mathfrak{S}_k$. Since the uniformity condition just stated (which is condition (3) of Lemma VIII) holds for s_n on each of the finite number of sets $\mathfrak{S}_1, \cdots, \mathfrak{S}_k$, it holds for the set $CE_k = \mathfrak{S}_1 + \cdots + \mathfrak{S}_k$. Any interval ω on CE_λ is on CE_k for some sufficiently great value of k. It then follows that the sequence $s_n = s_{n1}$ on $\mathfrak{S}_1$, $s_n = s_{n2}$ on $\mathfrak{S}_2$, $\cdots$ satisfies condition (3) of Lemma VIII on CE_λ.

If we start anew with E_λ, this process can be continued as in the situations above where transfinite induction is used, until a stage is reached at which a sequence s_n of summable functions is defined on R such that $s_n \to F'_s$ and $\int_\omega s_n\,dx\,dy \to F(\omega)$ uniformly in ω for ω on R. This completes the proof of Theorem XI.

This process for inverting the strong derivative of a function of intervals may be made the basis of a constructive definition of a non-absolutely convergent integral of a measurable function $f(x, y)$. A comparison with the definition of $I(\omega, f)$ given in the introduction reveals that the two definitions are the same. It is not possible to show that $I'_s = f$ almost everywhere on R, for this is not necessarily true, even when f is summable on R and $I(\omega, f) = \int_\omega f\,dx\,dy$. However, it is possible to replace F'_s by f in the proof of Theorem XI and thus arrive at the conclusion that there exists a sequence of summable functions on R with s_n tending to f and $\int_\omega s_n\,dx\,dy$ tending to $I(\omega)$ for every interval ω on R. In other words, the function $I(0, 0; x, y)$ is in class V_2 on R. It then follows that $I(\omega, f)$ satisfies all the conditions of Theorem IX, and we have the theorem:

If the measurable function $f(x, y)$ is such that $I(\omega, f)$ exists for every interval ω on R, then $I'(x, y)$ exists and is equal to $f(x, y)$ almost everywhere on R.

With the understanding that the definitions of functions in classes V_1 and V_2 have been stated in terms of intervals rather than in terms of the two variables x and y, we restate the descriptive definition of a non-absolutely convergent integral implied in Theorem IX.

Let the function $f(x, y)$ be measurable on R. If there exists a continuous additive function of intervals on R such that

(1) *$F'(x, y) = f(x, y)$ almost everywhere;*

(2) *$F(\omega)$ is in class V_2;*

(3) *if E is any closed set on R the points of E at which $F(\omega)$ is not of class V_1 and completely additive are non-dense on E;*

then $F(\omega)$ is the non-absolutely convergent integral of $f(x, y)$.

That $F(\omega)$ and $I(\omega, f)$ are identical was shown in the introduction, since it was there shown that any two functions satisfying the conditions of Theorem IX are identical.

We mention two questions we have not answered. Let $F(\omega)$ be a continuous additive function of intervals on R for which F' is finite at each point of R. Is there a process for inverting F' which is analogous to that given by Theorem X' for inverting F'_s? Is $F(\omega)$, or what is the same thing, $F(0, 0; x, y)$ in class V_2 on R?

All the results of this paper are valid in any number of dimensions. It is only necessary to replace two-dimensional intervals by n-dimensional intervals and to give a detailed description of $\mathcal{C}(R, E)$ for R an n-dimensional interval.

Acadia University and The University of Saskatchewan.

CAYLEY NUMBERS AND NORMAL SIMPLE LIE ALGEBRAS OF TYPE G

By N. Jacobson

In an earlier paper[1] we discussed the set $\mathfrak{D}(\mathfrak{A})$ of derivations in an arbitrary algebra $\mathfrak{A}$ (not necessarily associative), i.e., the operators D in $\mathfrak{A}$ such that

$$(x + y)D = xD + yD, \qquad (x\alpha)D = (xD)\alpha, \qquad (xy)D = x(yD) + (xD)y,$$

α being in the underlying field Φ. We noted that $\mathfrak{D}$ is closed with respect to addition, scalar multiplication and commutation $[D, E] \equiv DE - ED$. Hence $\mathfrak{D}$ is a Lie algebra over Φ. We shall show here that if $\mathfrak{A}$ is a generalized Cayley algebra and Φ is of characteristic 0, then $\mathfrak{D}$ is normal simple of type G and all such Lie algebras may be obtained in this way. The derivation algebras are isomorphic if and only if the Cayley systems are. If Φ is algebraically closed, these results have been indicated by Cartan.[2] The extension to the general case given here depends essentially on the determination of the automorphisms of $\mathfrak{D}$ in the algebraically closed case. The structure of Cayley systems has been obtained by Zorn.[3] We give several extensions of his theory.

We require below the theorem that if $\mathfrak{A}_P$ is the algebra obtained by extending Φ to P, then $\mathfrak{D}(\mathfrak{A}_P) = \mathfrak{D}_P$.[4] We note also that if S is either an automorphism or anti-automorphism in $\mathfrak{A}$ such that $(x\alpha)S = (xS)\alpha^s$, where $\alpha \rightarrow \alpha^s$ is an automorphism in Φ, then $S^{-1}DS$ is a derivation for every D in $\mathfrak{D}$. If $\alpha^s \equiv \alpha$, $D \rightarrow S^{-1}DS$ is an automorphism of $\mathfrak{D}$ over Φ.

1. Let Q be a (generalized) quaternion algebra over a field of characteristic $\neq 2$. We do not exclude the possibility that $Q = \Phi_2$, the 2-rowed matrix algebra. A (generalized) Cayley algebra $\mathfrak{A}$ is a vector space of order 2 over Q, $\mathfrak{A} = Q1 + Qe_4$, in which

$$(1) \qquad (a + be_4)(c + de_4) = (ac + \bar{d}b\alpha_4) + (da + b\bar{c})e_4,$$

where $\alpha_4 \neq 0$. If Q has basis $(1, e_1, e_2, e_3)$ such that $e_1^2 = \alpha_1$, $e_2^2 = \alpha_2$, $\alpha_i \neq 0$, $e_1e_2 = -e_2e_1 = e_3$ and $e_4^2 = \alpha_4$, then $1, e_1, \cdots, e_7$ is a basis for $\mathfrak{A}$ if $e_5 = e_1e_4$, $e_6 = e_4e_2$, $e_7 = e_3e_4$. $(1, e_1, e_4, e_5)$, $(1, e_4, e_2, e_6)$, $(1, e_3, e_4, e_7)$, $(1, e_1, e_6, e_7)$, $(1, e_2, e_5, e_7)$ and $(1, e_3, e_5, -e_6\alpha_1)$ are quaternion algebras. If e_i, e_j, e_k do not belong to one of these algebras, then $(e_ie_j)e_k = -e_i(e_je_k)$. It is sometimes

Received December 28, 1938; presented to the American Mathematical Society, December 28, 1938.

[1] *Abstract derivation and Lie algebras*, Trans. Amer. Math. Soc., vol. 42(1937), pp. 206–224.

[2] *Les groupes réels simples et continus*, Ann. de l'École Normale, vol. 31(1914), p. 298.

[3] *Alternativkörper und quadratische Systeme*, Hamb. Abhandl., vol. 9(1933), pp. 395–402.

[4] Loc. cit. (footnote 1), p. 213.

convenient to use

$$(2)\qquad (ae_4)b = (a\bar{b})e_4, \qquad a(be_4) = (ba)e_4, \qquad (ae_4)(be_4) = \bar{b}a\alpha_4$$

in place of (1).

Abstract characterizations of $\mathfrak{A}$ have been given by Zorn.[5] We recall the following properties. Any two elements of $\mathfrak{A}$ generate an associative algebra. If $x = a + be_4$, define $\bar{x} = \bar{a} - be_4$. Then the correspondence $x \to \bar{x}$ is an involutorial anti-automorphism in $\mathfrak{A}$. If $N(x) = x\bar{x} = \bar{x}x$, tr $(x) = x + \bar{x}$, these belong to Φ and $x^2 - \text{tr}\,(x)x + N(x) = 0$. We have $N(xy) = N(x)N(y)$, tr (x) is linear and tr xy = tr yx. Hence the function $(x, y) = \frac{1}{2}$ tr $x\bar{y}$ is a symmetric non-degenerate bilinear form whose corresponding quadratic form is $N(x)$.

We call the elements of trace 0 ($\bar{x} = -x$) vectors and let $\mathfrak{A}'$ denote their totality. Relative to (x, y) and $x \times y = \frac{1}{2}(xy - yx)$ we obtain a vector calculus whose fundamental properties are

$$(3)\qquad (x \times y, x \times y) = (x, x)(y, y) - (x, y)^2,$$

$$(4)\qquad (x \times y) \times z + x \times (y \times z) = 2y(x, z) - x(y, z) - z(x, y).$$

Equation (4) may be proved by noting the linearity and verifying for the basis e_i. $\mathfrak{A}$ is determined by the structure of $\mathfrak{A}'$ since $xy = -(x, y) + x \times y$ and hence

$$(\alpha + x)(\beta + y) = \alpha\beta + y\alpha + x\beta - (x, y) + x \times y.$$

Suppose f_1, f_2 are vectors such that $(f_1, f_1) \neq 0$, $(f_2, f_2) \neq 0$, $(f_1, f_2) = 0$. Let $f_3 = f_1 \times f_2$. Then by (3) and (4)

$$f_2 \times f_3 = f_1(f_2, f_2), \qquad f_3 \times f_1 = f_2(f_1, f_1),$$

$$(f_2, f_3) = (f_3, f_1) = 0, \qquad (f_3, f_3) = (f_1, f_1)(f_2, f_2).$$

We may choose an f_4 orthogonal ($(f_i, f_4) = 0$) to f_1, f_2, f_3. The vectors $f_5 = f_1 \times f_4$, $f_6 = f_4 \times f_2$, $f_7 = f_3 \times f_4$ together with f_1, f_2, f_3 are mutually orthogonal and their vector and scalar products are determined by (f_1, f_1), (f_2, f_2) and (f_4, f_4). $(1, f_1, \cdots, f_7)$ is a basis of the same type as $(1, e_1, \cdots, e_7)$ and $\mathfrak{A} = R + Rf_4$, $R = (1, f_1, f_2, f_3)$. R may be taken as any quaternion subalgebra of $\mathfrak{A}$. Hence in the original notation we may suppose that Q is any quaternion algebra in $\mathfrak{A}$.

In the generation (Q, e_4) we may replace e_4 by de_4. Then α_4 is replaced by $\alpha_4 N(d)$. Conversely, if $c + de_4$ satisfies the first two conditions in (2), we obtain by the linearity that c also satisfies them and hence $c = 0$. If $Q \cong \Phi_2$, any element is a norm in Φ_2 and hence we may suppose $\alpha_4 = 1$. If $(x, x) \neq 0$, x has an inverse. On the other hand, if there exists an $x \neq 0$ such that $(x, x) = 0$, either Q is matric or α_4 is a norm in Q, and we may then suppose $\alpha_4 = 1$. We readily obtain a matric subalgebra of $\mathfrak{A}$ and a normalized generation of the type described.

[5] Loc. cit. (footnote 3). Cf. also Zorn, *Theorie der alternativen Ringe*, Hamb. Abhandl. vol. 8(1931), pp. 123–147.

THEOREM. *Any two Cayley algebras which are not division algebras are isomorphic.*

If $\mathfrak{A}_1 = (e_0, \cdots, e_7)$ and $\mathfrak{A}_2 = (f_0, \cdots, f_7)$ are isomorphic Cayley algebras, the matrices (e_i, e_j), (f_i, f_j) are cogredient. Suppose conversely that these norm forms are equivalent; i.e., their matrices are cogredient. Choose e_1, e_2 in $\mathfrak{A}_1'$ such that $(e_1, e_2) = 0$, $(e_1, e_1) \neq 0$, $(e_2, e_2) \neq 0$ and f_1, f_2 in $\mathfrak{A}_2'$ such that $(f_i, f_j) = (e_i, e_j)$. By (3) and (4), $(1, e_1, e_2, e_3 = e_1 \times e_2)$ and $(1, f_1, f_2, f_3 = f_1 \times f_2)$ are isomorphic under the correspondence $e_i \rightarrow f_i$. If the other e's and f's are orthogonal to these, the matrices determined by them are cogredient.[6] Hence we may suppose that $(e_4, e_4) = (f_4, f_4)$ and then the correspondence $e_4 \rightarrow f_4$ will determine an extension of the isomorphism between $\mathfrak{A}_1$ and $\mathfrak{A}_2$.

THEOREM. *Any two Cayley algebras which have equivalent norm forms are isomorphic.*

THEOREM. *If Q and R are quaternion subalgebras of a Cayley algebra $\mathfrak{A}$, $e \rightarrow f = eS$ an isomorphism between them, then S may be extended to an automorphism in $\mathfrak{A}$.*

If Φ is extended to P, $\mathfrak{A}_P$ the extended algebra is evidently a Cayley algebra. Conversely, if $\mathfrak{A}_P$ is a Cayley algebra, $\mathfrak{A}$ has a non-degenerate form (x, y) defined in it. It follows that $\mathfrak{A}$ contains a quaternion subalgebra Q and an element e_4 orthogonal to Q, $(e_4, e_4) \neq 0$ and hence $\mathfrak{A}$ is a Cayley algebra.

2. Suppose D is a derivation in $\mathfrak{A}$. $1D = 0$ and hence (1) the set of multiples 1α is an invariant subspace relative to the derivation algebra $\mathfrak{D}$. Let $x \in \mathfrak{A}'$ and set $xD = x'$. Since $x^2 = -N(x)$, $xx' + x'x = 0$, $x(x')^2 = (x')^2x$. From $(x')^2 = x' \operatorname{tr}(x') - N(x')$, we obtain $xx' \operatorname{tr}(x') = 0$. If $xx' = 0$ and $x' \neq 0$, $x^2 = 0$. We may generate $\mathfrak{A}'$ by elements x such that $x^2 \neq 0$ and for these we obtain $\operatorname{tr} x' = 0$. It follows that this holds for every x in $\mathfrak{A}'$. Thus $\mathfrak{A}'$ is also invariant relative to $\mathfrak{D}$ and $\mathfrak{A}$ is mapped into $\mathfrak{A}'$ by $\mathfrak{D}$. If $x \in \mathfrak{A}'$, $\bar{x}D = -xD = \overline{xD}$. Since $\bar{1}D = \overline{1D}$, D commutes with the anti-automorphism $x \rightarrow \bar{x}$ in $\mathfrak{A}$.

If A is a linear transformation in the vector space $\mathfrak{A}'$, we define its adjoint as the linear transformation A^* such that $(x, yA) = (xA^*, y)$ for all x, y.[7] The correspondence $A \rightarrow A^*$ is an involutorial anti-automorphism in the algebra of linear transformations. Since $(x\bar{y})D \in \mathfrak{A}'$, $0 = \operatorname{tr}(x\bar{y})D = \operatorname{tr}(xD)\bar{y} + \operatorname{tr} x(\overline{yD}) = (xD, y) + (x, yD)$. Thus $D^* = -D$, i.e., every derivation is a skew-symmetric linear transformation.

If $\mathfrak{A} = (Q, e_4)$ and D is a derivation in Q,[8] it follows from (2) that we obtain all possible extensions of D to derivations in $\mathfrak{A}$ by setting $e_4D = ce_4$, c any

[6] This is proved by Witt, *Theorie der quadratischen Formen in beliebigen Körpern*, Jour. für Math., vol. 176(1936), p. 34.

[7] Cf. Jacobson, *Normal semi-linear transformations*, Amer. Jour. of Math., vol. 61(1939), p. 48.

[8] The derivations in Q are all inner, i.e., $xD = xd - dx$ for a fixed element d in Q. See Jacobson, loc. cit. (footnote 1), p. 215.

element in $Q' = \mathfrak{A}' \cap Q$. Any D is determined by its effect on e_1, e_2 and e_4. We shall show that D has the form

$$(5) \qquad e_1D = \sum_1^7 e_i\lambda_i, \qquad e_2D = \sum_1^7 e_i\mu_i, \qquad e_4D = \sum_1^7 e_i\nu_i,$$

where

$$(6) \quad \lambda_1 = \mu_2 = \nu_4 = \mu_1\alpha_1 + \lambda_2\alpha_2 = \nu_1\alpha_1 + \lambda_4\alpha_4 = \nu_2\alpha_2 + \mu_4\alpha_4$$
$$= \lambda_6\alpha_4\alpha_2 + \mu_5\alpha_1\alpha_4 - \nu_3\alpha_1\alpha_2 = 0.$$

Equations (6) are obtained from $e_i(e_iD) + (e_iD)e_i = 0$ $(i = 1, 2, 4)$, $\mathrm{tr}\,(e_1e_2)D = \mathrm{tr}\,(e_1e_4)D = \mathrm{tr}\,(e_2e_4)D = \mathrm{tr}\,(e_3e_4)D = 0$. Now suppose D satisfies (5) and define $e_3D = (e_1D)e_2 + e_1(e_2D)$, $e_5D = (e_1D)e_4 + e_1(e_4D)$, etc. By subtracting a suitable derivation obtained by extending a derivation in $(1, e_1, e_2, e_3)$ by $e_4 \to ce_4$, we obtain an E of the type (5) for which $\lambda_2 = \lambda_3 = \mu_1 = \mu_3 = \nu_5 = \nu_6 = \nu_7 = 0$. Similarly, if we use $(1, e_1, e_6, e_7)$ with e_4, we may obtain an F for which $\lambda_6 = \lambda_7 = \nu_2 = \nu_3 = 0$ also. Then by (6) $\mu_4 = \mu_5 = 0$. If we use $(1, e_1, e_4, e_5)$ with e_2 we obtain $\lambda_4 = \lambda_5 = \mu_6 = \mu_7 = \nu_1 = 0$. Thus D is a sum of derivations and hence is itself a derivation. D has order 14 over Φ.

We wish to show next that $\mathfrak{D}$ is an irreducible set of linear transformations in $\mathfrak{A}'$. For this purpose we consider the enveloping algebra $\mathfrak{R}$, i.e., the smallest algebra of linear transformations containing the operators of $\mathfrak{D}$ in $\mathfrak{A}'$. The derivations defined by

$$e_1D = e_2\alpha_1\lambda + e_3\mu, \qquad e_2D = -e_1\alpha_2\lambda - e_3\nu, \qquad e_4D = 0$$

have the matrices

$$\begin{pmatrix} L_1 & \\ & L_2 \end{pmatrix}$$

relative to the basis $(e_1, e_2, \cdots, e_7)$,[9] where

$$L_1 = \begin{pmatrix} 0 & -\lambda\alpha_2 & \mu\alpha_2 \\ \lambda\alpha_1 & 0 & -\nu\alpha_1 \\ \mu & -\nu & 0 \end{pmatrix}$$

and L_2 is a 4-rowed matrix. It is readily verified that the enveloping algebra of these matrices includes the matrices

$$\begin{pmatrix} A & \\ & B \end{pmatrix},$$

where A is an arbitrary 3-rowed matrix. Furthermore, if E is the derivation defined by $e_1E = e_2E = 0$, $e_4E = e_5$, then $\alpha_1^{-1}E^2$ has the matrix

$$\begin{pmatrix} 0 & \\ & 1 \end{pmatrix}.$$

[9] The matrix M of a linear transformation A relative to the basis $(e_1, e_2, \cdots, e_n)$ is defined by $(e_1A, e_2A, \cdots, e_nA) = (e_1, e_2, \cdots, e_n)M$.

Thus R contains the linear transformations whose matrices are

$$\begin{pmatrix} A & \\ & 0 \end{pmatrix} = \begin{pmatrix} A & \\ & B \end{pmatrix} - \begin{pmatrix} A & \\ & B \end{pmatrix}\begin{pmatrix} 0 & \\ & 1 \end{pmatrix}.$$

In particular $\mathfrak{R}$ contains the linear transformation E_{11} such that $e_1E_{11} = e_1$, $e_iE_{11} = 0$ and E_{12} such that $e_1E_{12} = e_2$, $e_iE_{12} = 0$ for $i \neq 1$. If we replace e_1, e_2 by the pair e_i, e_j, we see that $\mathfrak{R}$ contains E_{ii} and E_{ij} and hence $\mathfrak{R}$ is the complete set of linear transformations in $\mathfrak{A}'$. Evidently this implies that $\mathfrak{A}'$ contains no proper subspace invariant under all the transformations of $\mathfrak{D}$.

3. Assume from now on that Φ has characteristic 0. The irreducibility of $\mathfrak{D}$ implies that $\mathfrak{D} = \mathfrak{D}_1 \oplus \mathfrak{D}_2 \oplus \cdots \oplus \mathfrak{D}_r$ a direct sum of simple Lie algebras.[10] If D is any element of $\mathfrak{D}$, $D = \sum D_i$, D_i in $\mathfrak{D}_i$. The condition that two such elements be commutative is that their components D_i be commutative. It follows that the subalgebra $\mathfrak{C}(D)$ of derivations commutative with a fixed D is a direct sum of r subalgebras $\mathfrak{C}(D) \cap \mathfrak{D}_i \neq 0$. Now let D be the derivation defined by $e_1D = e_2\alpha_1$, $e_2D = -e_1\alpha_2$, $e_4D = 0$. Then

$$(e_1D, e_2D, \cdots, e_7D) = (e_2\alpha_1, -e_1\alpha_2, 0, 0, -e_6\alpha_1, e_5\alpha_2, 0)$$

and the conditions $e_1[D, E] = e_2[D, E] = e_4[D, E] = 0$ imply that

$$e_1E = e_2\lambda_2 + e_5\lambda_5 + e_6\lambda_6, \qquad e_2E = e_1\mu_1 + e_5\mu_5 + e_6\mu_6, \qquad e_4E = e_3\nu_3 + e_7\nu_7,$$

where

$$\mu_1\alpha_1 + \lambda_2\alpha_2 = \mu_5\alpha_1 - \lambda_6\alpha_2 = \mu_6 + \lambda_5 = \nu_3\alpha_1 - 2\lambda_6\alpha_4 = 0.$$

It follows that $\mathfrak{C}(D)$ has the basis D, E_1, E_2, E_3, where

$$(e_1E_1, e_2E_1, \cdots, e_7E_1) = (e_5, -e_6, -2e_7, 0, e_1\alpha_4, -e_2\alpha_4, -2e_3\alpha_4),$$

$$(e_1E_2, e_2E_2, \cdots, e_7E_2) = \left(e_6, e_5\frac{\alpha_2}{\alpha_1}, 2e_4\alpha_2, 2e_3\frac{\alpha_4}{\alpha_1}, e_2\alpha_4, e_1\frac{\alpha_4\alpha_2}{\alpha_1}, 0\right),$$

$$(e_1E_3, e_2E_3, \cdots, e_7E_3) = (e_2\alpha_1, -e_1\alpha_2, 0, 2e_7, e_6\alpha_1, -e_5\alpha_2, -2e_4\alpha_1\alpha_2),$$

whence

$$[E_1, E_2] = 2E_3\frac{\alpha_4}{\alpha_1}, \qquad [E_2, E_3] = -2E_1\alpha_2, \qquad [E_3, E_1] = -2E_2\alpha_1.$$

This table shows that the only ideals in $\mathfrak{C}(D)$ are (D) and (E_1, E_2, E_3) and $\mathfrak{C}(D) = (D) \oplus (E_1, E_2, E_3)$. Hence $r \leqq 2$ and if $r = 2$, $D \in \mathfrak{D}_1$ and $(E_1, E_2, E_3) \subset \mathfrak{D}_2$. The derivation $D' = \frac{1}{2}(D - E_3)$ satisfies $e_4D' = -e_7$, $e_7D' = e_4\alpha_1\alpha_2$, $e_1D' = 0$. Hence if we replace e_1, e_2, e_4 by e_4, e_7, e_1, the above argument shows that $D' \in \mathfrak{D}_1$ or $D' \in \mathfrak{D}_2$. Either case contradicts $\mathfrak{D}_1 \cap \mathfrak{D}_2 = 0$. Hence $r = 1$ and we have proved the following theorem.

[10] The irreducibility implies that $\mathfrak{D} = \mathfrak{S} \oplus \mathfrak{D}'$, where $\mathfrak{S}$ is a commutative algebra and $\mathfrak{D}'$ is semi-simple (Jacobson, *Rational methods in the theory of Lie algebras*, Annals of Math., vol. 36(1935), p. 877). By Cartan's theorem $\mathfrak{D}'$ is a direct sum of simple algebras (Thèse, p. 53). Evidently any commutative algebra is a direct sum of simple algebras of order 1.

THEOREM. *$\mathfrak{D}$ is a simple Lie algebra.*

Thus if Φ is algebraically closed, $\mathfrak{D}$ is the simple Lie algebra of order 14 in the Killing-Cartan list.[11] We may now apply Cartan's theory of representations.[12] The weights of any representation of $\mathfrak{D}$ have the form $m_1\lambda_1 + m_2\lambda_2 + m_3\lambda_3$, where $m_1 + m_2 + m_3 = 0$, $3m_i$ and $m_i - m_j$ are integers. If Λ is a weight, so are all the forms $\Lambda - \omega_i, \Lambda - 2\omega_i, \cdots, \Lambda - 3m_i\omega_i = -m_i\lambda_i + (m_i + m_j)\lambda_j + (m_i + m_k)\lambda_k$, $\omega_i = \frac{2}{3}\lambda_i - \frac{1}{3}\lambda_j - \frac{1}{3}\lambda_k$ and $\Lambda - \omega_{ij}, \Lambda - 2\omega_{ij}, \cdots, \Lambda - (m_i - m_j)\omega_{ij} = m_j\lambda_i + m_i\lambda_j + m_k\lambda_k$, $\omega_{ij} = \lambda_i - \lambda_j$. The last equation shows that if Λ is a weight, so are all the forms obtained by permuting the m's in Λ. The highest weight of an irreducible representation satisfies $m_1 \geqq m_2 \geqq m_3$, $m_2 \leqq 0$.

LEMMA 1. *The representation in $\mathfrak{A}'$ is the only one (in the sense of similarity) of order $\leqq 7$.*

Suppose Λ is the highest weight of such a representation. We distinguish three cases:

(1) $m_2 < 0$, $m_2 > m_3$. Here $|m_i| \neq |m_j|$, $i \neq j$ and hence we obtain at least 12 distinct weights.

(2) $m_2 = 0$. Then $\Lambda = k(\lambda_1 - \lambda_2)$ and k is an integer. Again we obtain more than 7 distinct weights.

(3) $m_2 < 0$ and $m_2 = m_3$. Here $\Lambda = k(\frac{2}{3}\lambda_1 - \frac{1}{3}\lambda_2 - \frac{1}{3}\lambda_3)$ and we obtain more than 7 distinct weights unless $k = 1$. In this case we obtain 6 distinct weights $\neq 0$ and the highest is $\frac{2}{3}\lambda_1 - \frac{1}{3}\lambda_2 - \frac{1}{3}\lambda_3$.

Since the highest weights of any two representations under consideration are equal, these representations are similar.

Let $D \to D^S = \bar{D}$ be an automorphism in $\mathfrak{D}$. This correspondence defines a second irreducible representation of $\mathfrak{D}$ in $\mathfrak{A}'$, and hence by Lemma 1 there exists a linear transformation S in $\mathfrak{A}'$ such that $\bar{D} = S^{-1}DS$ for all D restricted to $\mathfrak{A}'$. Since $D, \bar{D}$ are skew, $\bar{D} = S^*D(S^*)^{-1}$ and $SS^*D = DSS^*$. Since the enveloping algebra of $\mathfrak{D}$ is the complete set of linear transformations, $SS^* = 1\sigma$, $\sigma \neq 0$. We may replace S by $S_1 = S\tau$, $\tau^2 = \sigma^{-1}$ and obtain that S_1 is orthogonal, i.e., $(xS_1, yS_1) = (x, y)$. Define $1S_1 = 1$. Then $\bar{D} = S_1^{-1}DS_1$ in $\mathfrak{A}$ and S_1 is orthogonal. Let $(1, e_1, \cdots, e_7)$ be a basis as in §1 and $f_i = e_iS$. Then $(e_i, e_j) = (f_i, f_j)$ and hence $e_i^2 = f_i^2$. Let D be the derivation such that $(e_1D, e_2D, \cdots, e_7D) = (e_2\alpha_1, -e_1\alpha_2, 0, -e_7, -2e_6\alpha_1, 2e_5\alpha_2, e_4\alpha_1\alpha_2)$. Evidently the multiples of e_3 are the only elements in $\mathfrak{A}'$ annihilated by D, and hence the multiples of f_3 are the only elements annihilated by $\bar{D}$. Since

$$(f_1f_2)\bar{D} = (f_1\bar{D})f_2 + f_1(f_2\bar{D}) = f_2^2\alpha_1 - f_1^2\alpha_2 = 0,$$

$f_1f_2 = f_3\rho$, and since

$$(f_3, f_3) = (e_3, e_3) = -(e_1, e_1)(e_2, e_2) = -(f_1, f_1)(f_2, f_2) = (f_1f_2, f_1f_2),$$

[11] Cartan, Thèse, Paris, 1894, p. 94.

[12] Cf. Cartan, *Les groupes projectifs qui ne laissent invariante aucune multiplicité plane*, Bull. Soc. Math. de France, vol. 41(1913), pp. 53–96.

we have $\rho = \pm 1$, or $f_1 f_2 = \pm f_3$. Similarly $f_i f_j = f_k \rho_{ij} \epsilon_{ij}$, $\epsilon_{ij} = \pm 1$ if $e_i e_j = e_k \rho_{ij}$. By changing the sign of τ if necessary, we may suppose $f_1 f_2 = f_3$. Then $f_2 f_3 = f_1 \alpha_2$, $f_3 f_1 = f_2 \alpha_1$ and $e_1 \to f_1$, $e_2 \to f_2$, $e_4 \to f_4$ leads to an automorphism S_2 in $\mathfrak{A}$ such that $T = S_1 S_2^{-1}$ satisfies

$$1T = 1, \qquad e_i T = e_i \epsilon_i, \qquad \epsilon_i = \begin{cases} 1 & \text{if } i = 1, 2, 3, 4, \\ \pm 1 & \text{if } i = 5, 6, 7, \end{cases}$$

and $D_1 = T^{-1}DT \in \mathfrak{D}$ if D does. Using our special D, we obtain

$$\begin{aligned} e_5 D_1 &= e_5 T^{-1} DT = -2e_6 \epsilon_5 \epsilon_6 \alpha_1 \\ &= (e_1 e_4) D_1 = -e_6(\epsilon_7 + 1)\alpha_1, \end{aligned}$$

and hence $\epsilon_7 = 1$, $\epsilon_5 \epsilon_6 = 1$. If we use the derivation E such that $e_1 E = e_3$, $e_3 E = e_1 \alpha_2$, $e_4 E = e_6$ and compute $e_5 E_1$ in two ways, we obtain $\epsilon_6 = \epsilon_5 = 1$. Thus $T = 1$, $S_1 = S_2$.

Lemma 2. *If $D \to D^{\bar{s}}$ is an automorphism of $\mathfrak{D}(\mathfrak{A})$, $\mathfrak{A}$ the Cayley algebra over an algebraically closed field Φ, then there is a unique automorphism S in $\mathfrak{A}$ such that $D^{\bar{s}} = S^{-1}DS$.*

If S' and S'' both satisfy this condition, $1S' = 1S'' = 1$ and S' and S'' send $\mathfrak{A}'$ into itself. $(S')^{-1}S''$ commutes with all D in $\mathfrak{A}'$. Hence $S'' = \rho S'$. It follows immediately that $\rho = 1$.

4. Suppose $\mathfrak{A}_1 = (1, e_1, \cdots, e_7)$ and $\mathfrak{A}_2 = (1, f_1, \cdots, f_7)$ are two Cayley algebras over Φ of characteristic 0 such that $\mathfrak{D}(\mathfrak{A}_1) \cong \mathfrak{D}(\mathfrak{A}_2)$, where, say, $D \to E$ gives the isomorphism. If Ω is the algebraic closure of Φ, there is just one Cayley algebra $\mathfrak{A}$ over Ω and we may regard $\mathfrak{A}_1$ and $\mathfrak{A}_2$ as subrings of $\mathfrak{A}$ and $1, e_1, \cdots, e_7$ and $1, f_1, \cdots, f_7$ as bases for $\mathfrak{A}$ over Ω. Let $D_1, \cdots, D_{14}$ and $E_1, \cdots, E_{14}$ be corresponding bases for $\mathfrak{D}(\mathfrak{A}_1)$ and $\mathfrak{D}(\mathfrak{A}_2)$. Either of these sets forms a basis for $\mathfrak{D}(\mathfrak{A})$ over Ω. If

$$e_i D_k = \sum e_j \mu_{ji}^{(k)}, \qquad f_i E_k = \sum f_j \nu_{ji}^{(k)}, \qquad (i, j = 1, \cdots, 7; k = 1, \cdots, 14),$$

the matrices $\sum M^{(k)} \omega_k$ and $\sum N^{(k)} \omega_k$, $M^{(k)} = (\mu_{ij}^{(k)})$, $N^{(k)} = (\nu_{ij}^{(k)})$ correspond in different representations of $\mathfrak{D}(\mathfrak{A})$. Hence there exists a matrix Q with elements in Ω such that $Q^{-1}M^{(k)}Q = N^{(k)}$. Since the M's and N's have elements in Φ, we may suppose that Q has elements in Φ also and by choosing the basis $f_1, \cdots, f_7$ in $\mathfrak{A}_1'$ suitably, we may assume $M^{(k)} = N^{(k)}$. Let T be the linear transformation defined by $1T = 1$, $e_i T = f_i$. Then $D_k \to T^{-1} D_k T$ defines an automorphism in $\mathfrak{D}(\mathfrak{A})$. By the proof of Lemma 2, there exists a $\tau \neq 0$ such that $1, f_1\tau, \cdots, f_7\tau$ satisfies the same multiplication table as $1, e_1, \cdots, e_7$. Since $e_1 e_2 = e_3$, $f_1 f_2 \tau = f_3$ and $\tau \in \Phi$. Thus

$$S = \begin{cases} 1 & \text{in } (1), \\ T\tau & \text{in } \mathfrak{A}' \end{cases}$$

induces an isomorphism between $\mathfrak{A}_1$ over Φ and $\mathfrak{A}_2$ over Φ such that $E = S^{-1}DS$. S is unique since its extension is an automorphism in $\mathfrak{A}$ over Ω.

THEOREM. *If $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are Cayley algebras over Φ such that there exists an isomorphism $D \rightarrow E$ between their derivation algebras, then there exists a unique isomorphism S between $\mathfrak{A}_1$ and $\mathfrak{A}_2$ such that $E = S^{-1}DS$.*

THEOREM. *If $\mathfrak{A}$ is a Cayley algebra over Φ, $\mathfrak{D}$ its derivation algebra, then the group of automorphisms of $\mathfrak{D}$ is isomorphic to the group of automorphisms of $\mathfrak{A}$.*

If S is an automorphism in $\mathfrak{A}$, $D \rightarrow S^{-1}DS = D^{\bar{S}}$ is one in $\mathfrak{D}$. By the preceding theorem any $\bar{S}$ has this form and the corresponding S is unique.

5. A Lie algebra $\mathfrak{L}$ is of *type G* if $\mathfrak{L}_\Omega$ is isomorphic to the derivation algebra of the Cayley algebra over Ω. The latter is one of the five exceptional algebras in Cartan's list of the simple Lie algebras over Ω. If $\mathfrak{A}$ is any Cayley algebra over Φ, we have noted that $\mathfrak{D}_\Omega = \mathfrak{D}(\mathfrak{A}_\Omega)$ so that $\mathfrak{D}$ is of type G.

Now suppose $\mathfrak{L}$ is an arbitrary Lie algebra of type G. The elements of $\mathfrak{L}$ may be represented as derivations in $(1, e_1, \cdots, e_7)$ over Ω, where $\alpha_1 = \alpha_2 = \alpha_4 = 1$ in the notation of §2. Thus $\mathfrak{L} \cong$ the set $\sum_1^{14} D_i\gamma_i$, γ in Φ, and $\sum D_i\omega_i$, ω in Ω, give all the derivations in the algebra over Ω. The D's are determined as in (5) and the λ_i, μ_i, ν_i obtained by the derivations $\sum D_i\gamma_i$ generate a finite algebraic extension of Φ and may be taken as elements in a finite Galois extension P of Φ. Thus $\mathfrak{L}_\mathrm{P}$ is isomorphic to the Cayley algebra of $(1, e_1, \cdots, e_7)$ over P.

If $s \in \mathfrak{G}$ the Galois group of P over Φ, the correspondence $x = \sum_0^7 e_i\rho_i \rightarrow \sum e_i\rho_i^s = xS$ $(e_0 = 1)$ is an automorphism of $(1, e_1, \cdots, e_7)$ over P regarded as an algebra over Φ and $(x\rho)S = (xS)\rho^s$. It follows that $E_i = S^{-1}D_iS$ are linearly independent (over P) derivations and have the same multiplication table as the D_i. Hence $\sum D_i\sigma_i \rightarrow \sum E_i\sigma_i$, σ in P, is an automorphism in the derivation algebra and there exists an automorphism $\bar{S}$ of $(1, e_1, \cdots, e_7)$ over P such that $\sum E_i\sigma_i = \bar{S}^{-1}(\sum D_i\sigma_i)\bar{S}$. In particular, if $S_1 = S\bar{S}^{-1}$, $S_1^{-1}(\sum D_i\gamma_i)S_1 = \sum D_i\gamma_i$. Let $\mathfrak{A}$ be the subset of elements y such that $yS_1 = y$ for all s in $\mathfrak{G}$. $\mathfrak{A}$ is an algebra over Φ containing 1.

If $st = u$ in $\mathfrak{G}$, $ST = U$ and

$$\bar{U}^{-1}D_i\bar{U} = U^{-1}D_iU = T^{-1}S^{-1}D_iST = T^{-1}\bar{S}^{-1}D_i\bar{S}T$$
$$= T^{-1}\bar{S}^{-1}T(T^{-1}D_iT)T^{-1}\bar{S}T = (T^{-1}\bar{S}^{-1}T)\bar{T}^{-1}D_i\bar{T}(T^{-1}\bar{S}T).$$

$\bar{U}$ and $\bar{T}(T^{-1}\bar{S}T)$ are automorphisms in $(1, e_1, \cdots, e_7)$ over P and hence by the uniqueness noted above they are equal. If $U_1 = U\bar{U}^{-1}$, $T_1 = T\bar{T}^{-1}$, it follows that $S_1T_1 = U_1$. We require the

LEMMA. *Let $\mathfrak{R}$ be a vector space of order n over P, a separable Galois field over Φ. Suppose $s \rightarrow S_1$ is a (1-1) representation of the Galois group $\mathfrak{G}$ of P over Φ by semi-linear transformations such that $\mu S_1 = S_1\mu^s$. Then the order over Φ of $\mathfrak{R}_0$ the set of elements invariant under all S_1 is n and its extension $\mathfrak{R}_0\mathrm{P} = \mathfrak{R}$.*

The set of operators $\sum S_1\mu_s$, μ_s in P, is a cross product (P, S_1, 1) and hence is isomorphic to Φ_k the k-rowed matrix algebra over Φ. We may regard $\mathfrak{R}$ as a vector space over Φ. Then its order will be nk and the operators in (P, S_1, 1) are linear transformations and therefore determine matrices in Φ_{nk}. Thus we obtain a representation of Φ_k in Φ_{nk} such that the identities correspond. It is well known that this representation decomposes into n irreducible parts all equivalent to Φ_k. Hence we may write any x in $\mathfrak{R}$ uniquely as $x_1 + \cdots + x_n$, where the x_i form an invariant subspace $\mathfrak{R}^{(i)}$ relative to all $\sum S_1\mu_s$. The condition $xS_1 = x$ is equivalent to $x_iS_1 = x_i$ and therefore $\mathfrak{R}_0 = \mathfrak{R}_0^{(1)} + \cdots + \mathfrak{R}_0^{(n)}$, if $\mathfrak{R}_0^{(i)} = \mathfrak{R}^{(i)} \cap \mathfrak{R}_0$. Because of the similarity of the transformations induced by S_1 in $\mathfrak{R}^{(i)}$ and $\mathfrak{R}^{(j)}$ the order $(\mathfrak{R}_0^{(i)} : \Phi) = (\mathfrak{R}_0^{(j)} : \Phi)$ and hence $(\mathfrak{R}_0 : \Phi) = n(\mathfrak{R}_0^{(i)} : \Phi)$. Now if we consider the special case where $n = 1$ and the operators S_1 are the elements of the Galois group, we see that the elements left invariant by S_1 form a 1-dimensional subspace, namely, Φ itself.[13] It follows that $\mathfrak{R}_0$ has order n over Φ. If y_i is a vector in $\mathfrak{R}^{(i)}$ such that $y_iS_1 = y_i$, $\mathfrak{R}^{(i)}$ consists of the multiples $y_i\rho$ and hence $\mathfrak{R} = \mathfrak{R}_0\mathrm{P}$.

Returning to our special case, we see that $\mathfrak{A}$ has order 8 over Φ and $\mathfrak{A}\mathrm{P} = (1, e_1, \cdots, e_7)$ over P. Hence $\mathfrak{A}\mathrm{P} \cong \mathfrak{A}_\mathrm{P}$ and $\mathfrak{A}$ is a Cayley algebra. If $D = \sum D_i\gamma_i$, $(yD)S_1 = yD$, i.e., $yD \in \mathfrak{A}$ and hence $\sum D_i\gamma_i$ is a derivation in $\mathfrak{A}$. Since the D_i are independent, we obtain here all the derivations in $\mathfrak{A}$.

THEOREM. *A necessary and sufficient condition that a Lie algebra $\mathfrak{L}$ over Φ be of type G is that $\mathfrak{L} = \mathfrak{D}(\mathfrak{A})$, $\mathfrak{A}$ a Cayley algebra over Φ.*

6. The above results establish a complete equivalence between the problems of classifying and obtaining the automorphisms of Lie algebras of type G and the analogous problems for Cayley algebras. As was noted also the classification of Cayley algebras can be reduced to a question of equivalence of certain quadratic forms in eight variables. For certain special fields, this is readily accomplished. In particular, if Φ is real closed, there are two Cayley algebras and if Φ is an algebraic number field, there are 2^{r_1} such algebras, where r_1 is the number of real conjugate fields.

We remark finally that the arguments above are quite general. The special considerations are all contained in the study of the structure and automorphisms for the algebras over an algebraically closed field. We hope to apply this method to other types of Lie algebras in a later paper.

UNIVERSITY OF NORTH CAROLINA.

[13] This method may be used to prove that (P, S, 1) $\cong \Phi_k$. For we obtain here a realization of the cross product by linear transformations in P over Φ. Since the order of (P, S, 1) $= k^2$, this set includes all linear transformations and hence is isomorphic to Φ_k.

THE FIRST CANONICAL PENCIL

By P. O. Bell

I. Introduction

Among the most important covariant lines which lie in the tangent plane to a surface S at a point P_x are the first canonical edge of Green [5],[1] the first directrix of Wilczynski [7], the reciprocal [5] with respect to the surface S of the projective normal [4], and the reciprocal with respect to S of the axis of Čech [3]. In view of the fact that these covariant lines, each of which was discovered by a different author, were characterized by apparently unrelated properties, it has been considered remarkable that they all should pass through a common point of the tangent plane. This point has been called the *canonical point*. Wilczynski [7] and Green [5] have referred to lines in the tangent plane to S at P_x as lines of the *first kind*. Accordingly, a covariant line which passes through the *canonical point* has been called a *canonical line of the first kind*. The totality of canonical lines of the first kind form the *first canonical pencil* [1]. The primary purpose of the author in this note is to present a new geometric characterization of a general canonical line of the first kind. For this purpose the projective normal is first constructed in a new way.

II. The projective normal

Let the surface S be referred to its asymptotic net as parametric, and let us choose the associated fundamental differential equations in *Fubini's canonical form*

$$(1)\qquad \begin{cases} x_{uu} = px + \theta_u x_u + \beta x_v, \\ x_{vv} = qx + \gamma x_u + \theta_v x_v, \end{cases}$$

where $\theta = \log \beta\gamma$. Let l denote an arbitrarily chosen line of the first kind. The line l therefore intersects the u- and v-tangents to S at P_x in points ρ and σ whose general coördinates are of the forms $\rho = x_u - bx$, $\sigma = x_v - ax$, in which a and b are functions of u and v. Let l' denote the reciprocal of l with respect to S at P_x. Let τ and ω denote, respectively, the points distinct from P_x in which the line l' intersects the quadrics of Wilczynski and Lie [7] at the point P_x. The general coördinates of τ and ω may easily be found to be given by the expressions $\tau = z + (ab - \frac{1}{2}\theta_{uv})x$ and $\omega = \tau - \frac{1}{2}(\beta\gamma)x$, in which $z = x_{uv} - ax_u - bx_v$. For this purpose one would make use of the equations

$$(2)\qquad 2(x_2x_3 - x_1x_4) - \theta_{uv}x_4^2 = 0$$

Received December 15, 1938; presented to the American Mathematical Society, April 9, 1937.

[1] Numbers in brackets refer to the bibliography at the end of the paper.

and

$$2(x_2x_3 - x_1x_4) - (\beta\gamma + \theta_{uv})x_4^2 = 0 \tag{3}$$

for the quadrics of Wilczynski and Lie, respectively, referred to the reference tetrahedron whose vertices are the points x, x_u, x_v and x_{uv}. As the point P_x moves along a curve C_λ, defined by $dv - \lambda\, du = 0$ (λ is a function of u and v), the points τ and ω describe corresponding curves. The tangent lines at τ and ω to these curves intersect the tangent plane to S at P_x in points which we denote by T_λ and W_λ, respectively. The expression for the general coördinates of T_λ is given by a linear combination of τ and $\tau_u + \lambda\tau_v$ which does not contain x_{uv}. A similar combination of ω and $\omega_u + \lambda\omega_v$ gives the expression for the general coördinates of W_λ. The term of $\tau_u + \lambda\tau_v$ which involves x_{uv} is $(\theta_u + \lambda\theta_v - b - a\lambda)x_{uv}$; the same term appears in $\omega_u + \lambda\omega_v$. Hence, the expressions for T_λ and W_λ are

$$T_\lambda = \tau_u + \lambda\tau_v - (\theta_u + \lambda\theta_v - b - a\lambda)\tau,$$

$$W_\lambda = \omega_u + \lambda\omega_v - (\theta_u + \lambda\theta_v - b - a\lambda)\omega.$$

Making use of the expressions for τ and ω and the equations (1), we obtain

$$W_\lambda - T_\lambda = -\tfrac{1}{2}\beta\gamma[x_u + bx + \lambda(x_v + ax)]. \tag{4}$$

Let t_λ denote the line which is tangent to C_λ at P_x, let r denote the line joining W_λ and T_λ, and let ν_λ denote the point of intersection of t_λ and r. Since the right member of (4) is a linear combination of x and $x_u + \lambda x_v$ and is also a linear combination of the expressions for the coördinates of W_λ and T_λ, it is the expression for the coördinates of ν_λ. We shall call the point ν_λ the *ν-point* of t_λ, *corresponding to the point P_x and the line l.* Since the right member of (4) is a linear combination of $x_u + bx$ and $x_v + ax$, the point ν_λ, for any value of λ, lies on a straight line $\mathbf{l}$ which joins the points $\bar\rho$ and $\bar\sigma$ given by

$$\bar\rho = x_u + bx, \qquad \bar\sigma = x_v + ax.$$

We shall call the line $\mathbf{l}$ the *ν-associate* of the line l, corresponding to the point P_x of S. We state now

THEOREM 1. *As the direction λ is varied at P_x, the ν-point of t_λ, corresponding to the point P_x of S and the line l, describes the ν-associate $\mathbf{l}$ of the line l.*

The point μ of intersection of l and $\mathbf{l}$ has general coördinates of the form $\mu = ax_u - bx_v$. Since the points x_u and x_v lie on the line l_n, the reciprocal with respect to S of the projective normal, the point μ likewise lies on it. Hence we have

THEOREM 2. *The points μ_i ($i = 1, 2, \cdots$) which are the intersections of an arbitrarily chosen set of lines l_i of the first kind with the corresponding ν-associates $\mathbf{l}_i$ are collinear, and they determine the line l_n which is the reciprocal of the projective normal l_n'.*

Let t_i denote the tangents to S at P_x which pass through the points μ_i. The equations of the lines l_i, referred to the triangle of reference whose vertices are

the points x, x_u and x_v, are of the form $x_1 + b_ix_2 + a_ix_3 = 0$ $(i = 1, 2, \cdots)$. The equations for the corresponding ν-associates $\bar{l}_i$ and the tangents t_i are, respectively,

$$x_1 - b_ix_2 - a_ix_3 = 0, \qquad a_ix_3 + b_ix_2 = 0.$$

Therefore we obtain

THEOREM 3. *For an arbitrary value k of the set $(i = 1, 2, \cdots, k, \cdots)$ the harmonic conjugate of t_k with respect to the line l_k and its ν-associate $\bar{l}_k$ is the line l_n which is the reciprocal of the projective normal l'_n.*

Let t'_i denote the tangents to S at P_x which are conjugate, for corresponding values of i, to the tangents t_i at P_x. Now since for any value $i = k$ the tangent t'_k and the lines l'_k, $\bar{l}'_k$ and l_n, which are the reciprocals of l_k, $\bar{l}_k$ and l_n, are reciprocal polar lines of t_k, l_k, $\bar{l}_k$ and l_n with respect to the quadrics of Darboux, we have

THEOREM 4. *The lines t'_k, l'_k, $\bar{l}'_k$ and l'_n, for an arbitrary value $i = k$ of the set $(i = 1, 2, \cdots)$, are coplanar, and the harmonic conjugate of t'_k with respect to the lines l'_k and $\bar{l}'_k$ is the projective normal l'_n.*

Theorem 4 shows that the planes π_i, in which π_k is determined by l'_k and $\bar{l}'_k$, form a pencil whose axis is the projective normal.

III. The D_k curves of S

A curve of S will be called a *D_k curve* if at every point of it the curves of intersection of its asymptotic osculating quadrics of Bompiani [2] lie on the quadric of Darboux whose equation is [6]

$$2(x_2x_3 - x_1x_4) - [\beta\gamma(1 - k) + \theta_{uv}]x_4^2 = 0, \qquad k = \text{const.} \tag{5}$$

For each value of k there are just three D_k curves of S through P_x.

The curvilinear differential equation for these curves is obtained as follows. The equations for the asymptotic osculating quadrics Q_v and Q_u of a curve C_λ at P_x are

$$\begin{aligned} &2(x_2x_3 - x_1x_4) - 2\gamma\lambda x_4(x_3 - \lambda x_2) \\ &\qquad + \{\gamma[-\lambda' - \gamma\lambda^3 + (\psi - \theta_v)\lambda^2 - (2\varphi - \theta_u)\lambda] - (\beta\gamma + \theta_{uv})\}x_4^2 = 0, \end{aligned} \tag{6}$$

and

$$\begin{aligned} &2\lambda^3(x_2x_3 - x_1x_4) + 2\beta\lambda x_4(x_3 - \lambda x_2) \\ &\qquad + \{\beta[\lambda' - \beta + (\varphi - \theta_u)\lambda - (2\psi - \theta_v)\lambda^2] - (\beta\gamma + \theta_{uv})\lambda^3\}x_4^2 = 0, \end{aligned} \tag{7}$$

respectively, where φ and ψ are defined by

$$\varphi = (\log \beta\gamma^2)_u, \qquad \psi = (\log \beta^2\gamma)_v.$$

On multiplying both members of (6) by β and both members of (7) by γ and adding corresponding members of the resulting equations, we obtain the equation

$$\begin{aligned} &2(\gamma\lambda^3 + \beta)(x_2x_3 - x_1x_4) \\ &\qquad - (\beta\theta_{uv} + 2\gamma\beta^2 + \gamma\beta\varphi\lambda + \gamma\beta\psi\lambda^2 + [2\gamma^2\beta + \gamma\theta_{uv}]\lambda^3)x_4^2 = 0, \end{aligned} \tag{8}$$

which is the equation of a quadric of Darboux if $\gamma\lambda^3 + \beta \neq 0$. It is equivalent to (5) if, and only if, λ is chosen so that

$$\gamma\lambda^3 + (\psi/[1 + k])\lambda^2 + (\varphi/[1 + k])\lambda + \beta = 0. \tag{9}$$

The curves of intersection of Q_u and Q_v which correspond to C_λ at P_x cannot lie on more than one quadric of Darboux at P_x of C_λ, for a unique quadric of Darboux passes through a given point not in the tangent plane to S at P_x. We therefore have the curvilinear differential equation for the D_k curves of S in the form

$$\gamma\, dv^3 + (\psi/[1 + k])\, dv^2\, du + (\varphi/[1 + k])\, dv\, du^2 + \beta\, du^3 = 0, \quad k = \text{const.} \tag{10}$$

IV. The geometric characterization of a general canonical line of the first kind

There are ∞^3 quadrics which have contact of the second order with S at a point P_x. The equation of a general one of these can be written in the form

$$x_2x_3 - x_1x_4 + x_4(k_2x_2 + k_3x_3 + k_4x_4) = 0, \tag{11}$$

where k_2, k_3, k_4 are arbitrary functions of u and v as P_x moves over S, and are constants when P_x remains fixed. The quadric (11) cuts the surface S in a curve which has a triple point at P_x and whose triple point tangents are in the directions which satisfy the equation

$$\gamma\, dv^3 + 3k_3\, dv^2\, du + 3k_2\, dv\, du^2 + \beta\, du^3 = 0, \qquad k_2,\ k_3 \text{ arbitrary.} \tag{12}$$

Since equation (10) is a special case of equation (12), we have

THEOREM 5. *There are ∞^1 quadrics which have second order contact with S at P_x and which intersect S in a curve whose triple point tangents are the tangents of the D_k curves of S at P_x.*

We shall call these the *D_k quadrics at P_x*. The equation of a general one of them has the form

$$x_2x_3 - x_1x_4 + (\tfrac{1}{3}\varphi/[1 + k])x_2x_4 + (\tfrac{1}{3}\psi/[1 + k])x_3x_4 + k_4x_4^2 = 0, \tag{13}$$

k = constant, k_4 = an arbitrary function of u, v. The polar line of the projective normal with respect to the quadric (13) is the line whose equations are

$$x_4 = 0, \qquad x_1 - (\tfrac{1}{3}\varphi/[1 + k])x_2 - (\tfrac{1}{3}\psi/[1 + k])x_3 = 0, \qquad k = \text{const.} \tag{14}$$

These are the equations for a general canonical line of the first kind. We may, therefore, state

THEOREM 6. *The polar line of the projective normal with respect to a D_k quadric at P_x is a canonical line of the first kind which is independent of the choice of the quadric. By a proper selection of the constant k this line may be made to become any desired line of the first canonical pencil.*

Equations (14) show that if $k = -\frac{2}{3}$, the line is the first directrix of Wilczynski; if $k = -2$, it is the reciprocal of the axis of Čech with respect to the surface S; if $k = -\frac{1}{3}$, it is the first canonical edge of Green; and if $k = \infty$, it is the reciprocal of the projective normal with respect to the surface S.

Bibliography

1. P. O. Bell, *Covariant configurations related to analytic curved surfaces*, this Journal, vol. 4(1938), pp. 590–599.
2. E. Bompiani, *Ancora sulla geometria delle superficie considerate nello spazio rigato*, Reale Accademia dei Lincei, Rome, Rendiconti, (6), vol. 4(1926), pp. 262–267.
3. E. Čech, *L'intorno di un punto d'una superficie considerato dal punto di vista proiettivo*, Annali di Matematica Pura ed Applicata, (3), vol. 31(1922), pp. 191–206.
4. G. Fubini, *Fondamenti della geometria proiettivo-differenziale di una superficie*, Reale Accademia delle Scienze, Torino, Atti, vol. 53(1918), pp. 1032–1043. See also G. M. Green, Bulletin of the American Mathematical Society, vol. 23(1916), pp. 73–74, Abstract.
5. G. M. Green, *Memoir on the general theory of surfaces and rectilinear congruences*, Transactions of the American Mathematical Society, vol. 20(1919), pp. 79–153.
6. E. P. Lane, *The correspondence between the tangent plane of a surface and its point of contact*, American Journal of Mathematics, vol. 48(1926), pp. 204–214.
7. E. J. Wilczynski, *Projective differential geometry of curved surfaces* (Memoirs 2-3), Transactions of the American Mathematical Society, vol. 9(1908), pp. 79–120; 293–315.

University of Kansas.

THE NON-EXISTENCE OF A CERTAIN TYPE OF CONTINUOUS TRANSFORMATION

By O. G. Harrold, Jr.

1. It has been shown recently[1] that there exist continuous transformations defined on an arc A such that, if B denotes the image space, each point in B has *exactly* k inverse points in A ($k = 3, 4, 5, \cdots$). Thus for $k = 1, 3, 4, 5, \cdots$ it is possible to define an exactly $(k, 1)$ continuous transformation[2] on an arc. For $k = 3, 4, 5, \cdots$ the image may be taken to be a circle. We show that for $k = 2$ the circle cannot be an exactly $(k, 1)$ image of an arc, and in fact, no exactly $(2, 1)$ continuous transformation can be defined on an arc. This is the same as saying that it is impossible for any Jordan continuum to be so generated as the path of a moving point, whose coördinates are continuous functions of the time $(0 \leqq t \leqq 1)$, that every point of the continuum is passed through exactly twice.

2. In this section it will be shown that no exactly $(2, 1)$ continuous transformation exists carrying an arc into either an arc or a circle. First, we establish

Lemma A. *Let T be a continuous transformation of A into B such that each point $b \in B$ has at most two inverse points in A. If $T(xy) = ab$ preserves end-points, where xy is an arc in A and ab is an arc in B, then T is topological on xy.*

This lemma can be proved immediately if we notice that it is essentially equivalent to the following theorem concerning a real continuous function:

If $f(x)$ is continuous for $0 \leqq x \leqq 1$, $f(0) = 0$, $f(1) = 1$, $0 \leqq f(x) \leqq 1$, and finally for each y $(0 \leqq y \leqq 1)$ $f^{-1}(y)$ consists of one or two values, then f is monotonic on $0 \leqq x \leqq 1$.

From Lemma A it follows easily that our image space can be no arc. Evi-

Received January 4, 1939; in revised form, May 31, 1939.

[1] The following example, formulated by G. E. Schweigert, shows that an exactly $(3, 1)$ continuous transformation can be defined on an arc, the image being a circle.

Let C be a circle and (y_n) a sequence of distinct points on C in cyclic order with $\lim y_n = y_1$. Let the arc of C between y_n and y_{n+1} be denoted by Y_n. Set $x_n = 1 - n^{-1}$. Let the arc of the unit interval between x_n and x_{n+1} be denoted by X_n. Define T as follows. Map X_1 topologically on Y_1 with $T(x_1) = y_1$. For $n \geqq 1$, map X_{2n+1} on $Y_n + Y_{n+1}$ topologically with $T(x_{2n+1}) = y_n$. For $n \geqq 1$, map X_{2n} on Y_n with $T(x_{2n}) = y_{n+1}$. $T(\lim x_n) = y_1$.

If we add to the unit interval the semi-closed interval $1 < x \leqq 2$ and map it in one-to-one continuous fashion on the whole circle so that $T(1) = T(2)$, an exactly $(4, 1)$ continuous transformation is obtained. The method obviously admits extension.

[2] By a $(k, 1)$ transformation is understood a single-valued transformation such that every point in the image space has at most k inverse points. An exactly $(k, 1)$ transformation is a single-valued transformation such that every point in the image space has exactly k inverse points.

dently, all that remains to be shown is that an arc xy cannot be transformed in an exactly (2, 1) continuous fashion onto an arc ab in such a way that $(x + y)T^{-1}(a) = 0$. Supposing such a transformation were possible, let $T^{-1}(a) = a^1 + a^2$, where a^1 precedes a^2. Suppose $T(a^1a^2) = aj$. Let $j^1 \epsilon a^1a^2 \cdot T^{-1}(j)$. Then each of a^1j^1 and j^1a^2 contains an inverse to an arbitrary inner point of aj. Hence $T(xa^1 - a^1)$ does not have a as a limit point, and the continuity of T is denied.

Next let us assume that $T(xy)$ is a circle C. Let u^1, u^2, v^1, and v^2 be distinct points on the arc xy and a and b distinct points of C such that (1) $T(u^1) = T(u^2) = a$, (2) $T(v^1) = T(v^2) = b$, (3) u^1 is x and either u^2 or v^2 is y. Let s^1, s^2, and s^3 be the three components of $xy - (u^1 + u^2 + v^1 + v^2)$. Let d and e denote points of C which are separated by a and b, so that adb and aeb are arcs whose sum is C. For each i $(i = 1, 2, 3)$ $T(s^i)$ is a subset (proper or improper) of one of the open arcs adb, aeb. Thus one of these arcs, say adb, contains the image under T of just one of the segments s^i, say s^1. Then $T(\bar{s}^1) = adb$, and from this it follows that s^1 has end-points u^i and v^j. Then by Lemma A, T is topological on $\bar{s}^1$ and thus $T^{-1}(d)$ is a single point. This completes the proof that neither an arc nor a circle can be the image of an arc under an exactly (2, 1) continuous transformation.

3. In this section it is shown that, if T is an exactly (2, 1) continuous transformation defined on an arc A, $B = T(A)$ can contain no triod $ab + ac + ad$ in which two of the arcs, say ab and ac, are free arcs. Three cases may be distinguished according as $T^{-1}(a) = a^1 + a^2$ contains two, one, or no end-points of the arc A. As the method involved is the same in all three cases, we shall consider only one of them. The last one involves the longest argument; hence we consider that one.

Let the end-points of A be x and y. If xa^1, say, contains inverses of each point of a sequence of points on ab converging to a, let $x^1 \epsilon xa^1$ be sufficiently near a^1 so that $T(x^1a^1) \subset ab$, and, hence is an arc, aj. Let x^2 be the first point on x^1a^1 as it is traversed from a^1 which belongs to $T^{-1}(j)$. By Lemma A, T is topological on a^1x^2. The set of inverses to ab not in a^1x^2 must have a^1 or a^2 (or both) for a limit point. In either event another arc, a^2y^2 say, $T(a^2y^2) \subset ab$, can be determined on which T is topological. Similarly, corresponding to the free arc ac there are determined two arcs in A each of which has a point of $T^{-1}(a)$ as an end-point and such that T is topological on each one. Hence a^1 and a^2 can each be enclosed in an open set U^i which contains no inverses to $ad - a$, but the continuity of T implies that $T^{-1}(a)$ contains at least one more point, and this is not possible.

4. The image space B must be stably regular (bestandig regular).[3] To establish this it is sufficient to show that for an arbitrary continuum $K \subset B$, K contains a free arc of B. We prove first the weaker property that each arc t

[3] A compact continuum is called stably regular provided the enclosure of its set of ramification points is totally disconnected.

in B contains a free arc. To show this suppose, on the contrary, each point of t is a limit point of $B - t$. Either $T^{-1}(t)$ is totally disconnected, or, since $T^{-1}(t) \subset A$, $T^{-1}(t)$ contains an arc u. The first possibility is ruled out by

LEMMA B. *The dimensionality of a compact metric space P is unaltered by an exactly $(k, 1)$ continuous transformation.*

This lemma is a consequence of the following theorems of Hurewicz[4] on dimension theory:

(1) If B is the continuous image of the compact metric space A, and if $\dim B < \dim A$, then some point x in B is such that $T^{-1}(x)$ contains a continuum.

(2) If B is the continuous image of the compact metric space A, and if B_k is the set of points of B with exactly k inversive images, $\dim B_k \leqq \dim A$.

If $T^{-1}(t)$ contains an arc u, set $T(u) = t^1$. Let the end-points of t^1 be c^1 and d^1. At least one component of $u - T^{-1}(c^1 + d^1)u$ has an enclosure u^1, one of whose end-points, x, maps into c^1 and the other, y, into d^1, hence by Lemma A, T is topological on the arc u^1. Set $h = T^{-1}(t^1) - u^1 + (x + y)$. If h is totally disconnected, a contradiction arises, for if p is an inner point in t^1 which has two inverses in h (surely such a point exists, since h is totally disconnected), then p has three inverses, since it already has one in u^1. Suppose, on the other hand, h contains an arc v. Then $T(v) = t^2 \subset t^1$. Let $v^1 \subset v$ be an arc mapping topologically into t^2. Let p be an inner point of t^2 and (y_n) a sequence of points in $B - t$ converging to p. Then $T^{-1}(y_n) \subset A - (u^1 + v^1)$. Hence p must have an inverse in the enclosure of $A - (u^1 + v^1)$, and thus p has at least three inverse points. We have shown that every arc in B contains a free arc of B. This implies that every continuum K in B contains a free arc of B. For $T^{-1}(K)$ cannot be totally disconnected by Lemma B and hence contains an arc u. But $T(u)$ is a compact locally connected continuum in K and thus contains an arc t, which by the above remarks must contain an arc which is free in B. Thus B must be stably regular.

5. In view of the preceding we have only to consider the case in which B is a stably regular curve in which no two maximal free arcs abut. It will be shown that, if such a transformation were possible, it would be possible to define an auxiliary transformation carrying an arc into a curve containing no free arc by an exactly $(2, 1)$ transformation.

Corresponding to a stably regular curve B in which no two maximal free arcs abut an upper semi-continuous decomposition can be effected as follows. The points of the hyperspace C are the free arcs of B and the points of B belonging to no free arc. Clearly, this satisfies the requirements of an upper semi-continuous decomposition. Let the corresponding continuous transformation of B into C be S, $S(B) = C$. It is evident that, if $x \epsilon C$, $S^{-1}(x)$ is a point or an arc, hence S is monotone on B. In fact, if B^1 is any continuum in B which contains all free arcs of which it contains both end-points, S is monotone on B^1.

[4] W. Hurewicz, Proceedings of the Amsterdam Academy, vol. 29–30.

By virtue of the fact that an end-point of a curve is a limit point of cut points of the curve, it is easy to show that B contains at most one end-point. In view of this, it is next shown that each maximal free arc u in B has at least one end-point which is a limit point of a sequence of distinct topological circles in B. Let p be an end-point of u that is not an end-point of B. If p were not a limit point of simple closed curves in B, there would exist arbitrarily small connected neighborhoods in B containing p whose enclosures would be acyclic curves. Now let qp be an arc in such a curve with only the point p in u. Since u is the unique free arc in B containing p, the enclosure of the set of ramification points of B contains points on $qp - p$ arbitrarily near p. We may thus determine a sequence of mutually exclusive arcs in this acyclic curve each of which has but one point on qp. Since in any hereditarily locally connected continuum an arbitrary subset has only a finite number of components of diameter greater than any given positive number, B must have end-points arbitrarily near p, and this is not possible.

Suppose now that the curve C contains a free arc u. Let u^1 be an arc entirely in the interior of u. Let V be an open set in u containing u^1. Since B is compact and S is monotone, $U = S^{-1}(u^1)$ is a continuum. Since B is stably regular, U is a continuous curve and contains a free arc. By the remarks of the preceding paragraph, one or the other of the end-points of this free arc must be a limit point of topological circles in B. Let e be a topological circle in the open set $W = S^{-1}(V)$, $U \subset W$. The set $S(e) = f$ is a non-degenerate continuum in V, i.e., an arc. Adjoin to e all free arcs j in B both of whose end-points lie in e. The resulting set e^1 is clearly a continuum in the stably regular curve B which is cyclicly connected. Moreover, $S(e^1) = f$ and S is monotone on e^1. Let three points in order on the arc f be a^1, a^2, and a^3. The set $S^{-1}(a^2)$ is a free arc or a point in the cyclic curve e^1, and hence does not separate $S^{-1}(a^1)$ and $S^{-1}(a^3)$. Hence S is not non-alternating on e^1 and thus not monotone.[5] Hence C contains no free arc.

Define $R(A) = ST(A) = C$. Then R is a continuous transformation of the arc A into the curve C which contains no free arcs. We shall show that to every point $x \in C$, $R^{-1}(x)$ consists of two components. First, let x be a point in C whose inverse in B is a point x^1. Since T is an exactly (2, 1) transformation, clearly x has two inverse points in A. Next, let x be a point in C whose inverse in B is a free arc x^2. The inverse of x^2 in A, by virtue of Lemma B, cannot be totally disconnected and hence contains an arc j, which is a component of $T^{-1}(x^2)$. Clearly, since T is continuous and x^2 is a maximal free arc in B, the end-points of j must map into one or both of the end-points of x^2. The inverse of each end-point of x^2 must contain one point in each non-degenerate component of $T^{-1}(x^2)$, or, both inverses of an end-point of x^2 lie in the same component. Hence there must be at least one and at most two non-degenerate components in $T^{-1}(x^2)$.

[5] G. T. Whyburn, *Non-alternating transformations*, American Journal of Mathematics, vol. 56(1934), pp. 294–302.

In the second case there can be no degenerate components and in the first only one, hence in all cases a point $x \epsilon C$ has two and only two inverse components in A. Factor the transformation $R = T_2T_1(A)$, where $T_1(A) = A^1$, the transformation T_1 being monotone and T_2 light.[6] Since the monotone image of an arc is an arc, the continuous transformation T_2 carries the arc A^1 into the curve C containing no free arcs in such a way that every point in C has exactly two inverse points in A^1. This denies the result of §4. Thus we have completed the proof that no exactly (2, 1) continuous transformation can be defined on an arc.

OREGON STATE COLLEGE.

[6] Whyburn, loc. cit., Theorem 3.2.

FORMAL POWER SERIES TRANSFORMATIONS

By Daniel C. Lewis, Jr.

1. **Introduction and summary of principal results.** Consider an analytic non-singular transformation T of the neighborhood of the origin of an n-dimensional space into itself. The t-th iterate of T, denoted by T^t, is defined when t is an integer and may be represented by convergent power series whose coefficients are functions of t. Furthermore the relation $T^{t+\tau} = T^t T^\tau$ holds for all integral values of t and τ. The question which arises is this: Is it possible to define T^t for non-integral values of t so that this relation holds for all t and τ? If we require the dependence on t to be analytic, the answer is certainly in the negative (save for comparatively rare exceptions), as can be shown by examples ($n = 2$).[1] It is, however, in general, possible, as we prove in this paper, to define the coefficients of the series for T^t for all values of t, without regard to whether or not they converge, in such a manner that $T^{t+\tau} = T^t T^\tau$ holds in a purely formal sense. Furthermore, the coefficients may be defined as comparatively simple functions of t, being, in fact, polynomials in t and a finite number of expressions of the form $e^{\mu t}$. This result is important for dynamical theory, but we do not restrict ourselves here to the type of transformation arising in dynamics. The special properties of the dynamical transformations will be studied in a future paper.

Previous work in this field has been done by C. L. Bouton,[2] who treated the case when the matrix of coefficients of the linear terms is the unit matrix, and by G. D. Birkhoff,[3] who treated the so-called "conservative" surface transformations ($n = 2$) which arise in dynamics.

The case when T is linear is of special importance. Here questions of convergence have no place. The problem is essentially that of defining the t-th

Received February 20, 1939.

[1] If such convergent series exist, they must satisfy a system of differential equations of the form $dx_i/dt = X_i(x)$ in which the X_i are convergent series in $x_1, \cdots, x_n$ but independent of t. Cf. §6.2. In the case of "conservative" surface transformations ($n = 2$) these equations have an analytic first integral and are therefore completely integrable. But examples can be constructed of conservative transformations of "stable type" which in this sense cannot be integrable. Cf., for example, G. D. Birkhoff, *Dynamical Systems*, American Mathematical Society Colloquium Publications, vol. 9, New York, 1927, p. 250. The general solutions of the "non-integrable" equations there discussed yield such transformations when the independent variable takes on the value 2π. For further information on this subject cf. G. D. Birkhoff, *Surface transformations and their dynamical applications*, Acta Mathematica, vol. 43(1922), pp. 1–119, especially p. 16.

[2] Bulletin of the American Mathematical Society, vol. 23(1916), p. 73.

[3] Acta Mathematica, loc. cit. The restriction to conservative transformations is not entirely essential.

power of a matrix for all t, the solution of which is known.[4] The solution given in §3 is, however, obtained by entirely different methods and in a form which is necessary for our development of the theory of the non-linear case. On account of the intrinsic interest of the linear case, we have devoted somewhat more space to its treatment than would otherwise seem desirable. In this connection it should be observed that the results for the non-linear case can also be interpreted in terms of pure algebra. Our results are completely independent of whether or not the series defining T converge or not; and the set of all formal non-singular transformations T form a group with respect to the usual product relation.[5]

The heart of the paper is in §4 where we obtain our most affirmative results for transformations of so-called "pseudo-incommensurable type". This type of transformation is to be regarded as the general type in a very wide sense. In §5 it is shown that the situation for those comparatively rare transformations that are not of this general type is essentially different. It is nevertheless also shown that, if T is *any* formal non-singular transformation, there can always be found a positive integer k such that T^k is of the pseudo-incommensurable type.

Our work is concerned for the most part with the complex domain, although at the end of §§3 and 6 we consider briefly the situation for real transformations.

2. **Preliminary lemmas and definitions.**

The following assortment of lemmas, mostly of an elementary nature, is collected here for convenience:

LEMMA I. *If* $\sum_{m,n} A_{mn} e^{\Lambda_m t} t^n$ *represents a finite sum extended over certain non-negative integral values of* m *and* n, *if* $\Lambda_m \not\equiv \Lambda_{m'}$ (mod $2\pi i$) *unless* $m = m'$, *and if the sum vanishes for all sufficiently large integral values of* t, *then all the coefficients* A_{mn} *must vanish.*

Proof. Denote the above sum by $f(t)$. Let M denote the maximum real part of any of the Λ's and let N denote the largest of the n's accompanying terms having Λ's whose real parts are each equal to M. Then $f(t)e^{-Mt}t^{-N}$ will have the form $\sum_k B_k e^{i\theta_k t} + o(t)$ as $t \to \infty$, where the $i\theta$'s are some of the imaginary parts of those Λ's whose real parts are equal to M, and are consequently not congruent to each other (mod $2\pi i$). The B's are those A's accompanying the corresponding terms in the original sum. It follows that $\sum_k B_k e^{i\theta_k(t+\tau)} = o(t)$ for large integral values of t and $\tau = 0, 1, 2, \cdots$. We regard these equations as linear equations with the B's (h, in number, say) as the unknowns. In virtue of the fact that the θ's are not mutually congruent it is found that the determinant

[4] Cf. J. H. M. Wedderburn, *Lectures on Matrices*, Amer. Math. Soc. Colloquium Publications, vol. 17, New York, 1934, pp. 119–123.

[5] For the rôle played by formal power series in algebra cf. Saunders Mac Lane, *Some recent advances in algebra*, Amer. Math. Monthly, vol. 46(1939), pp. 3–19, where reference is made to the fundamental work of S. Lefschetz.

of the first h of these equations (i.e., the equations obtained by taking $\tau = 0, 1, 2, \cdots, h-1$) is bounded away from zero as $t \to \infty$. In fact, this determinant, up to a factor of absolute value one, turns out to be the Vandermonde determinant formed from $e^{i\theta_1}, \cdots, e^{i\theta_h}$. It follows that $B_k = o(t)$, which means, since B_k is independent of t, that $B_k = 0$. A repetition of this process proves that all the coefficients must vanish.

LEMMA II. *Let $\xi_1, \cdots, \xi_s$ denote any s numbers such that no linear combination of them with integral coefficients is a non-vanishing integral multiple of $2\pi i$. Let $P(t)$ represent a polynomial in the following $2s+1$ expressions:*[6]

$$t, \quad e^{+\xi_1 t}, \quad \cdots, \quad e^{+\xi_s t}, \quad e^{-\xi_1 t}, \quad \cdots, \quad e^{-\xi_s t}.$$

Then, if $P(t)$ vanishes for all sufficiently great positive integral values of t, it must vanish identically.

Proof. $P(t)$ can be written in the form of a finite sum of the type considered in Lemma I, where the Λ's are linear combinations of the ξ's with integral coefficients. No two of these Λ's can be congruent to each other modulo $2\pi i$ without being actually equal to each other; otherwise we could find a linear combination of the Λ's with integral coefficients equal to a non-vanishing integral multiple of $2\pi i$, contrary to hypothesis. Two or more terms of the type $Ae^{\Lambda t}t^n$ with numerically equal Λ's and n's are merged into one, so that all the hypotheses of Lemma I are fulfilled. The present lemma therefore follows as a corollary of the preceding.

The difference equation $\Delta f(t) \equiv f(t+1) - f(t) = e^{\xi t}$ ($\xi \not\equiv 0$, mod $2\pi i$) together with the initial condition $f(0) = 0$ is well known and is readily verified to admit the solution

$$f(t) = f_0(\xi, k, t) \equiv \frac{e^{\xi t} - e^{2k\pi i t}}{e^{\xi} - 1},$$

where k is an arbitrary integer. By repeated differentiation with respect to ξ we find that the equation $\Delta f(t) = e^{\xi t}t^n$ together with the same initial condition is satisfied by

$$f(t) = f_m(\xi, k, t) \equiv \frac{\partial^m}{\partial \xi^m} f_0(\xi, k, t).$$

Now $f_0(\xi, k, t)$ has a removable singularity for $\xi = 2k\pi i$, and the Taylor series development is easily seen to be of the following form:

$$f_0(\xi, k, t) = e^{2k\pi i t} \sum_{\nu=1}^{\infty} \frac{L_\nu(t)}{\nu!} (\xi - 2k\pi i)^{\nu-1},$$

[6] In the future we shall, for brevity, refer to such a function as a polynomial in the expressions $e^{\pm\xi_k t}t^l$.

where $L_\nu(t)$ is a real polynomial in t lacking a constant term.[7] It follows that $f_m(\xi, k, t)$ also has a removable singularity for $\xi = 2k\pi i$ and that

$$f_m(2k\pi i, k, t) \equiv e^{2k\pi i t} \frac{L_{m+1}(t)}{m+1} \qquad (k = 0, \pm 1, \pm 2, \cdots)$$

satisfies the difference equation $\Delta f(t) = e^{2k\pi i t} t^m$.

We now introduce, for convenience, the "chief sum" of a function $f(t)$, denoted by $Sf(t)$. We proceed to define $Sf(t)$ for certain simple types of functions in the following three steps:

I. $Se^{\xi t} t^m = f_m(\xi, 0, t)$, where m is a non-negative integer and $\xi \not\equiv 0 \pmod{2\pi i}$.

II. $$Se^{2k\pi i t} t^m = f_m(2k\pi i, k, t) \equiv e^{2k\pi i t} \frac{L_{m+1}(t)}{m+1} \qquad (k = 0, \pm 1, \pm 2, \cdots).$$

III. $S[f_1(t) + f_2(t)] = Sf_1(t) + Sf_2(t)$ and $Skf(t) = kSf(t)$, where $k = \text{const}$.

The following lemmas are now obvious from the preceding discussion:

LEMMA III. *The chief sum is now well defined for all functions $f(t)$ which can be written as polynomials in a finite number of expressions of the type $e^{\pm\mu t} t^l$, and $Sf(t)$ is another function of the same type. It vanishes for $t = 0$ and satisfies the difference equation $\Delta Sf(t) = f(t)$.*

LEMMA IV. *If a polynomial $f(t)$ in expressions of the type $e^{\pm\mu t} t^l$ takes on only real values, when t is real, then $Sf(t)$ also is real when t is real.*

Notice that to prove Lemma IV it is not sufficient to remark that, if $\Delta[x(t) + iy(t)] = f(t) + ig(t)$, where x, y, f, g, and t are real, then $\Delta x(t) = f(t)$ and $\Delta y(t) = g(t)$. For $g(t) \equiv 0$ does not imply that $y(t)$ is constant but only that it is periodic with period 1. The actual proof depends in a sufficiently obvious way on DeMoivre's theorem and an examination of the explicit formulas available for evaluating the chief sum.

In addition to the notation already mentioned, we introduce here the following conventions to be used in the sequel:

A point $(x_1, \cdots, x_n)$ in n-dimensional complex space will be denoted simply by x.

Square matrices with n rows will always be represented by capital letters; their elements by the corresponding small letters with subscripts to denote their rows and columns. The one exception to this rule is the unit matrix I whose elements are denoted by δ_{ij} according to the usual notation. The rule is, however, not to be interpreted as meaning that all capital letters represent matrices.

[7] It is interesting (but unessential) to note the well-known fact that $L_\nu(t) = B_\nu(t) - B_\nu(0)$, where $B_\nu(t)$ is the Bernoulli polynomial of degree ν.

3. **The linear case.** Let us consider a non-singular linear transformation L of the form

$$x_i(1) = \sum_{j=1}^{n} a_{ij} x_j(0) \qquad (i = 1, \cdots, n), \tag{1}$$

which sends the point $x(0)$ into the point $x(1)$. It is our first purpose to find a square matrix $Y(t)$ depending analytically on t in such a manner that, when t is a non-negative integer, $Y(t)$ coincides with A^t, the matrix of coefficients of L^t. In other words, if L^t sends the point $x(0)$ into $x(t)$, we require that

$$x_i(t) = \sum_{j=1}^{n} y_{ij}(t) x_j(0), \tag{2}$$

at least for non-negative integral values of t. Since

$$x_i(t+1) = \sum_j a_{ij} x_j(t) = \sum_{j,k} a_{ij} y_{jk}(t) x_k(0) = \sum_k y_{ik}(t+1) x_k(0),$$

it is clear that $y_{ik}(t)$ for each value of k must satisfy the linear system of difference equations

$$y_i(t+1) = \sum_{j=1}^{n} a_{ij} y_j(t) \tag{3}$$

and the initial conditions

$$y_i(0) = \delta_{ik}, \tag{4}$$

at least so far as non-negative integral values of t are concerned. We immediately add, however, the stronger requirement that the $y_{ik}(t)$ satisfy (3) and (4) for all t. This requirement certainly suffices for the validity of (2) for integral values of t. From the well-known theory of systems of difference equations with constant coefficients[8] we are able to satisfy (3) and (4) by means of a set of functions $y_{1k}(t), \cdots, y_{nk}(t)$ depending linearly on n functions of the type $\lambda_h^t t^{l_h}$, where $\lambda_1, \cdots, \lambda_r$ are the r distinct roots ($r \leqq n$) of the equation $\det(A - \lambda I) = 0$ and l_h takes on non-negative integral values less than the multiplicity of the root λ_h. Here λ_h^t is regarded as single valued, being interpreted as meaning $e^{\mu_h t}$, where μ_h is a fixed determination of $\log \lambda_h$. These determinations once made, the solution of (3) and (4) in this form is unique.

We have also the matrix relation $Y(t + \tau) = Y(t)Y(\tau)$ which is known to hold for all non-negative integral values of t and τ. With the aid of Lemma I,

[8] I do not know of any adequate treatment in the literature of this subject. Nevertheless the theory is well known, the situation being analogous to the corresponding situation for differential equations. The best reference I know of is to N. E. Nörlund, *Differenzenrechnung*, Berlin, 1924, pp. 295–298. Here the case of a single difference equation of order m (say) in a single unknown is exhaustively treated. The system of n equations (3) can be reduced to one, or more, such equations by forming differences and elimination of $n - 1$, or less, of the unknowns.

we prove that *this relation holds for all values of t and τ*. This matrix equation stands in fact for the n^2 ordinary equations

$$f_{ij}(t, \tau) \equiv y_{ij}(t + \tau) - \sum_{k=1}^{n} y_{ik}(t)y_{kj}(\tau) = 0 \qquad (i, j = 1, \cdots, n),$$

which are known to hold for non-negative integral values of t and τ. Furthermore $f_{ij}(t, \tau)$ considered as a function of t alone can be written in the form of a finite sum of the type considered in Lemma I. Hence it must vanish whenever τ is a non-negative integer, regardless of the value of t. Now consider $f_{ij}(t, \tau)$ as a function of τ alone. A second application of Lemma I yields the desired result, that $f_{ij}(t, \tau)$ vanishes for every value of t and τ.

Thus we have succeeded in defining L^t in the form (2) for all values of t in such a way that $L^{t+\tau} = L^t L^\tau$ and so that it agrees with the ordinary definition when t is an integer.

It is not by any means claimed that this definition of L^t (or A^t) for non-integral values of t is unique. But, if $\bar{Y}(t)$ is another matrix of solutions of (3) and (4), it must be related to $Y(t)$ by a matrix equation of the form $\bar{Y}(t) = Y(t)P(t)$, where $P(t)$ is periodic in t with period 1 and reduces to the unit matrix I when $t = 0$; and, if the relation $\bar{Y}(t)\bar{Y}(\tau) = \bar{Y}(t + \tau)$ is to hold, we must have

$$Y(t)P(t)Y(\tau)P(\tau) = Y(t + \tau)P(t + \tau). \tag{5}$$

It is not important for our purpose to discuss criteria for the existence of non-trivial matrices $P(t)$ which satisfy these conditions, but we content ourselves by showing with the help of a simple example that in certain cases such periodic matrices exist in great profusion. In fact, let A be a diagonal matrix ($n = 2$), both elements in the main diagonal being equal to e^μ. Then $Y(t)$ is also a diagonal matrix with both non-vanishing elements equal to $e^{\mu t}$. It commutes with every matrix and hence the condition (5) reduces to $P(t)P(\tau) = P(t + \tau)$. A matrix $P(t)$ satisfying all our conditions is now readily written down, namely,

$$P(t) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e^{2k\pi i t} & 0 \\ 0 & e^{2l\pi i t} \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1},$$

where k and l are any two integers and $ad - bc \neq 0$.[9] Thus A^t in this case has an infinite number of determinations because of the arbitrariness in the choice of a, b, c and d, as well as of k and l.

It is also interesting to remark that the infinitesimal transformation belonging to the continuous group L^t, defined above, is easily obtained. The essential facts may be formulated as follows: Consider the system of linear differential equations with constant coefficients

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} b_{ij}x_j, \qquad b_{ij} = \left.\frac{dy_{ij}}{dt}\right|_{t=0} \qquad (i, j = 1, \cdots, n).$$

[9] For another discussion of this subject from an entirely different point of view, cf. J. H. M. Wedderburn, loc. cit.

The solution of this differential system which takes on given initial values, $x_1(0), \cdots, x_n(0)$, *is given by the right members of equations* (2), which we now regard as defining the transformation L^t for all t.

The proof consists in differentiating the following obvious equations (which merely express the fact that $L^{t+\tau} = L^t L^\tau$) with respect to τ and then setting $\tau = 0$:

$$x_i(t + \tau) = \sum_{j=1}^{n} y_{ij}(\tau) x_j(t) \qquad (i = 1, \cdots, n).$$

If the original transformation L is real, it is not possible in general to choose μ_h ($= \log \lambda_h$) in such a way that the transformation L^t is real for non-integral real values of t. *If, however, none of the real roots of* $\det(A - \lambda I) = 0$ *is negative, it is always possible to choose the μ's so that L^t is real for all real t.* This is because the complex roots occur in conjugate imaginary pairs and the corresponding logarithms can also be chosen in conjugate imaginary pairs, so that the solution of (3) and (4) will be real, involving possibly trigonometric functions as well as real exponential functions in a well-known manner. It is understood, of course, in this connection that the logarithms of the positive roots are to be taken real. A necessary and sufficient condition that $\det(A - \lambda I) = 0$ have no negative real roots is obviously that L be the square of a real linear transformation.

4. **The non-linear case with special reference to "incommensurable exponents".** We consider a formal transformation T of the form

$$(6) \qquad \begin{aligned} x_i(1) = f_i[x(0)] &\equiv \sum_{j=1}^{n} a_{ij} x_j(0) \\ &+ \sum_{\mu=2}^{\infty} \left\{ \sum_{\alpha_1 + \cdots + \alpha_n = \mu} a_{i\alpha_1 \cdots \alpha_n} x_1^{\alpha_1}(0) \cdots x_n^{\alpha_n}(0) \right\} \quad (i = 1, 2, \cdots, n). \end{aligned}$$

We assume that the determinant $|a_{ij}|$ of coefficients of the linear terms is not zero. It is not necessary to assume the convergence of (6). The iterate T^t, where t is an integer, may be formed according to the usual formal rules for working with power series. Suppose it is written in the form

$$(7) \qquad \begin{aligned} x_i(t) = \varphi_i[t, x(0)] &\equiv \sum_{j=1}^{n} y_{ij}(t) x_j(0) \\ &+ \sum_{\mu=2}^{\infty} \left\{ \sum_{\alpha_1 + \cdots + \alpha_n = \mu} y_{i\alpha_1 \cdots \alpha_n}(t) x_1^{\alpha_1}(0) \cdots x_n^{\alpha_n}(0) \right\}. \end{aligned}$$

If the series (6) converge for $x(0)$ sufficiently close to the origin, the same is true for (7). But this fact does not greatly interest us. Here again we wish to interpolate, so that the $y(t)$'s, which are at first defined only for integral values of t, are defined and are analytic for all values of t. We furthermore desire the relations $T^{t+\tau} = T^t T^\tau$, which are known to hold for integral values of t and τ, to continue to hold, at least in a formal sense, for all t and τ.

It is clear that the $y_{ij}(t)$ may be determined exactly as in §3. We determine the other y's by induction as follows: Suppose all the coefficients of terms of degree less than N in (7) have already been determined as polynomials in the expressions $\lambda_k^{\pm t} t^l$. We shall show how the coefficients of degree N are also determined as such polynomials. From the relations $\varphi_i[t + 1, x(0)] \equiv f_i[\varphi[t, x(0)]]$, we find on equating coefficients that the functions $y_{1\alpha_1\cdots\alpha_n}(t), \cdots, y_{n\alpha_1\cdots\alpha_n}(t)$ must satisfy (for each set $\alpha_1, \cdots, \alpha_n$ whose sum is $N > 1$) a system of difference equations of the form

$$y_i(t + 1) = \sum_{j=1}^{n} a_{ij} y_j(t) + P_i(t), \tag{8}$$

together with the initial conditions

$$y_i(0) = 0 \qquad (i = 1, \cdots, n). \tag{9}$$

Here the $P_i(t)$ are polynomials in other y's, which are coefficients of terms of degree less than N. It follows from our inductive hypothesis that $P_i(t)$ is a known polynomial in the $\lambda_k^{\pm t} t^l$.

We now solve (8) by a well-known method analogous to Lagrange's procedure of variation of constants in the theory of differential equations. We know by §3 that the $y_{ik}(t)$ form a linearly independent set of solutions for the corresponding homogeneous system (3). Let us now seek to satisfy (8) by functions of the form

$$y_i(t) = \sum_{k=1}^{n} u_k(t) y_{ik}(t) \qquad (i = 1, \cdots, n). \tag{10}$$

Substituting in (8) we find that these equations may be written in the form

$$\sum_{k=1}^{n} [u_k(t + 1) - u_k(t)] y_{ik}(t + 1) + \sum_{k=1}^{n} u_k(t) y_{ik}(t + 1)$$
$$= \sum_{k=1}^{n} u_k(t) \sum_{j=1}^{n} a_{ij} y_{jk}(t) + P_i(t).$$

But, in virtue of (3), the last sum on the left side cancels with the sum on the right side. Hence we can satisfy (8), if we can find functions $u_1(t), \cdots, u_n(t)$ satisfying

$$\sum_{k=1}^{n} y_{ik}(t + 1) \Delta u_k(t) = P_i(t) \qquad (i = 1, \cdots, n).$$

Now we know from §3 that the matrix $Y(t + 1)$ has an inverse $Y(-t - 1)$. Hence

$$\Delta u_k(t) = Q_k(t), \tag{11}$$

where $Q_k(t)$ is also a polynomial in the $\lambda_k^{\pm t} t^l$. It follows from Lemma III that we have only to take $u_k(t) = SQ_k(t)$ to complete our inductive definition of the y's as polynomials in the $\lambda_k^{\pm t} t^l$.

At this point we must restrict attention to the case of "pseudo-incommensurable exponents", which is said to obtain when there exist numbers $\xi_1, \cdots, \xi_s$ with the following two properties:

I. *Each of the μ's ($\mu_h = \log \lambda_h$) is expressible in the form $\mu_h = \sum_{i=1}^{s} N_{hi}\xi_i$, where the N's are integers, positive, negative, or zero.* They may even all be zero for a given h.

II. *No linear combination of the ξ's with integral coefficients is a non-vanishing integral multiple of $2\pi i$.*

The case $s = 0$ is not excluded. In this case the set of ξ's is vacuous and all the μ's are zero.

The formal condition, $T^{t+\tau} = T^tT^\tau$, is equivalent to infinitely many conditions of the form $f(t, \tau) = 0$, where f is a polynomial in a finite number of the $y(t)$'s, $y(\tau)$'s and $y(t + \tau)$'s. Hence f is a polynomial in the $e^{\pm\mu t}t^l$ and the $e^{\pm\mu\tau}\tau^l$; and therefore also it may be considered as a polynomial in the $e^{\pm\xi t}t^l$ and $e^{\pm\xi\tau}\tau^l$. Now $f(t, \tau)$ is known to vanish when t and τ are integers. Applying Lemma II, we conclude that it vanishes when τ is an integer and t is arbitrary. A second application of the lemma, if we regard $f(t, \tau)$ as a function of τ and hold t fast at an arbitrary value, shows that $f(t, \tau)$ vanishes for all t and τ, as we wished to prove.

5. **The case of commensurable exponents.** Our proof that $T^{t+\tau} = T^tT^\tau$ breaks down, of course, if we do not make the hypothesis of pseudo-incommensurability introduced above. We first show by means of an example that the commensurable case is essentially different. Namely, let $n = 1$; let T be given by

$$x(1) = -x(0) + [x(0)]^2,$$

and let T^t be given by

$$x(t) = \sum_{h=1}^{\infty} y_h(t)[x(0)]^h.$$

If we choose $\log(-1) = \pi i$, we find according to the scheme of §§3 and 4 that $y_1(t) = e^{\pi i t}$, $y_2(t) = \frac{1}{2}(e^{2\pi i t} - e^{\pi i t})$, and $y_3(t) = -e^{3\pi i t}t - \frac{1}{2}e^{2\pi i t} + \frac{1}{2}e^{\pi i t}$. Among the conditions to be satisfied, if $T^{t+\tau} = T^tT^\tau$, is the following:

$$y_3(t + \tau) = y_1(\tau)y_3(t) + 2y_2(\tau)y_1(t)y_2(t) + y_3(\tau)[y_1(t)]^3.$$

It is readily verified that this condition is *not* satisfied for all values of t and τ by the above explicit expressions for y_1, y_2, y_3.

Whether or not it would be possible to make another determination of the y's which would satisfy our infinitely many conditions for the case of commensurable exponents, I do not know. For the present we must content ourselves in proving that *if T is an arbitrary formal transformation of the type* (6), *there exists another formal transformation S which is a positive integral power of T,*

say $S = T^k$, *such that, for a suitable choice of the exponents, S is of the pseudo-incommensurable type.* We can then, by the methods of §4, define S^t for all values of t in such a way that $S^{t+\tau} = S^t S^\tau$; but it will not in general happen that $S^{1/k}$ will be T.

Proof. Suppose that there exist s ($\leqq r$) independent relations of the form

$$\sum_{\beta=1}^{r} p_{\alpha\beta}\mu_\beta = 2K_\alpha \pi i \qquad (\alpha = 1, 2, \cdots, s) \tag{12}$$

(where the p's and K's are integers) such that any further relation of this type is linearly dependent on these. If no such relations exist, T itself is of pseudo-incommensurable type, since in that case we could take $\mu_\beta = \xi_\beta$. It is easily seen that there exist $\nu = r - s$ ($\geqq 0$) numbers $\xi_1, \cdots, \xi_\nu$ and integers k_α, $P_{\alpha\beta}$, and M_α such that

$$k_\alpha \mu_\alpha = \sum_{\beta=1}^{\nu} P_{\alpha\beta}\xi_\beta + 2M_\alpha \pi i \qquad (\alpha = 1, \cdots, r; k_\alpha > 0), \tag{13}$$

nor can there be any linear relation connecting the ξ's and $2\pi i$ with integral coefficients not all zero, since otherwise we could eliminate at least one of the ξ's from (13) after multiplying each of the r equations (13) by suitable positive integers, and this would imply the existence of at least $s + 1$ independent relations of the type (12), contrary to hypothesis.

The case $\nu = 0$, in which the set of ξ's is vacuous, is not excluded. This means that each of the μ's is separately commensurable with $2\pi i$ in the usual sense.

Let $k = k_1 k_2 \cdots k_r$ and take $S = T^k$. The matrix of coefficients of the linear terms of S will be A^k and the roots of the equation det $(A^k - \lambda I) = 0$ will be precisely the k-th power of the roots of det $(A - \lambda I) = 0$. One determination of the corresponding logarithms will be $k\mu_\alpha$ which, in virtue of (13), is a linear combination with integral coefficients of the ξ's and $2\pi i$. Another determination of these logarithms is obtained by dropping the integral multiples of $2\pi i$ from the aforementioned linear combinations. With this last determination of the logarithms of the roots of det $(A^k - \lambda I) = 0$, we see that S is of the pseudo-incommensurable type, as defined in §4.

6. Concluding remarks.

6.1. *A uniqueness theorem.* We learned in §4 how the coefficients $y(t)$ of (7) could be determined as polynomials in expressions of the form $e^{\pm\mu t}t^l$ ($l = 0, 1, 2, \cdots$; μ_h = a particular determination of log λ_h). It can be proved immediately with the help of Lemma II that in the pseudo-incommensurable case this is the only determination of this type possible. For, let $\bar{y}(t)$ be a second determination of one of the coefficients as a polynomial in expressions of the form $e^{\pm\mu t}t^l$. Then $y(t) - \bar{y}(t) = 0$ whenever t is an integer. Now the left side of this equation can be considered as a polynomial in expressions of the form $e^{\pm\xi t}t^l$, where no linear combination of the ξ's is a positive multiple of $2\pi i$. It follows from Lemma II that $y(t) - \bar{y}(t) \equiv 0$.

6.2. *The formal infinitesimal transformation.* The condition $T^{t+\tau} = T^t T^\tau$ may be written in the form

$$x_i(t + \tau) = \varphi_i[\tau, x(t)] \qquad (i = 1, \cdots, n), \tag{14}$$

where $x_i(t)$ stands for the formal series $\varphi_i[t, x(0)]$ in $x_1(0), \cdots, x_n(0)$ and (14) is to be interpreted as a formal identity in these variables. Differentiating with respect to τ and then setting $\tau = 0$, we obtain

$$\frac{dx_i}{dt} = \left.\frac{\partial \varphi_i[\tau, x(t)]}{\partial \tau}\right|_{\tau=0}.$$

Hence, denoting by $U_i(x)$ the formal power series $\frac{\partial}{\partial \tau} \varphi_i[\tau, x]_{\tau=0}$, we have the following theorem:

In the pseudo-incommensurable case, in which the coefficients $y(t)$ have been determined as in §4, the right members of (7) satisfy the formal system of differential equations

$$\frac{dx_i}{dt} = U_i(x)$$

and the initial conditions $x_i = x_i(0)$ when $t = 0$.

6.3. *Invariant formal series, etc.* A formal power series $W(x)$ in the variables $x_1, \cdots, x_n$ is said to be invariant under (7), if

$$W[x(t)] \equiv W[x(0)] \tag{15}$$

holds formally as an identity in $x_1(0), \cdots, x_n(0)$ when t is 1, and hence when t is any positive integer. The theorem, that this identity must therefore hold for all t in the pseudo-incommensurable case, is a special case of the following theorem:

If the formal series $x_i(t) = \varphi_i[t, x(0)]$ and a finite number of their partial derivatives with respect to the $x(0)$'s satisfy a formal relation of the type

$$\Omega\left[x(0), \varphi(t), \frac{\partial \varphi}{\partial x_1}, \frac{\partial \varphi}{\partial x_2}, \cdots, \frac{\partial^p \varphi}{\partial x_n^p}\right] = 0 \tag{16}$$

for positive integral values of t, it must also satisfy this relation for all values of t, at least, if T is of the pseudo-incommensurable type.

Here $\Omega(x, z, z^{(1)}, z^{(2)}, \cdots, z^{(pn)})$ denotes a formal power series in $x_1, \cdots, x_n$, $z_1, \cdots, z_n, z_1^{(1)}, \cdots, z_n^{(1)}$, etc.

The proof consists in the remark that (16) stands for an infinite number of polynomial relations in the y's which hold when t is a positive integer, and hence by Lemma II they must hold for all t.

It also follows from §6.2 that a necessary and sufficient condition that $W[x]$ be invariant under (7) in the pseudo-incommensurable case is that formally:

$$\sum_{i=1}^{n} \frac{\partial W}{\partial x_i} U_i = 0. \tag{17}$$

6.4. *Real transformations.* If the transformation T is the square of a real transformation, so that the matrix A has no latent roots which lie on the negative axis of reals, we have already remarked in §3 that μ_h $(= \log \lambda_h \; (h = 1, \cdots, r))$ may be chosen in such a way that the matrix $Y(t)$ must be real for real t. It follows from Lemma IV that the inductive definition of the other coefficients $y(t)$ given in §4 leads also to real values. The choice of the μ's which leads to this real determination of the $y(t)$'s is such that those μ's which are not real occur in conjugate imaginary pairs. A real formal transformation, *with which are associated μ's determined in this way*, is hereafter called briefly a "real transformation". Let those μ's which are not real be denoted by $\sigma_\nu \pm i\omega_\nu$ $(\nu = 1, \cdots, \rho \leqq \frac{1}{2}r)$. In order that a "real transformation" should be of pseudo-incommensurable type it is obviously sufficient, though not necessary, that the ω's be representable in the form

$$\omega_\nu = \sum_{i=1}^{s'} N_{\nu i}\theta_i,$$

where the N's are integers and the θ's are real numbers such that no integral linear combination of them is a non-vanishing integral multiple of 2π. If this condition is not satisfied for a given "real transformation" T, it can be shown that it will be satisfied for a suitable integral power of T, at least, if the ω's of the new "real transformation" T^k are suitably chosen. The proof is merely a slight modification of the proof appearing in §5, in which we dealt with ξ's and μ's instead of θ's and ω's. For the transformation T^k, which is of pseudo-incommensurable type, all the results of §§6.1, 6.2, and 6.3, of course, automatically hold.

CORNELL UNIVERSITY.

TOPOLOGICAL AND METRICAL POINTS OF VIEW IN THE THEORY OF SETS AND FUNCTIONS OF REAL VARIABLES

By A. Denjoy

The theory of sets and functions of real variables finds its most important applications to the natural sciences in connection with those functions which are solutions of ordinary differential equations, partial differential equations, or integral equations. These solutions, while regular in the region R where they satisfy the equations in question, generally cease to be regular on the boundary of R. This boundary may result from the nature of the functions themselves, or it may coincide with the locus along which the determining conditions of the solution are supposed given. Moreover, these latter conditions may be incompatible with the regularity of the solution on the boundary of R. Whatever may be the case, the region of validity of the solution is limited by a boundary along which the resolving function ceases to satisfy the given equation and along which it generally loses its character of analyticity or even of continuity. The function considered is often completely characterized by its behavior in the neighborhood of the boundary. That behavior is determined at each point by numbers presenting themselves as limits of expressions which are continuous away from the boundary, but undefined on the boundary itself. These limits, being variable with the point considered on the boundary, are a priori functions of real variables of a fairly general nature, not necessarily analytic, and usually discontinuous. Their study is therefore dependent upon the general theory of functions of real variables. But a fundamental remark is that these characteristic functions present themselves as limits, greatest limits, lowest limits, or merely accumulation values of certain families of functions which depend on parameters, and vary continuously with each of these parameters as long as the parameters remain positively far from the respective limits to which they are tending.

The most interesting questions to be set will often consist of seeking the properties of size (that is, of linear, quadratic, or higher order measure) of the sets E lying on the boundary F, where the numbers characterized above satisfy diverse conditions set a priori.

In this kind of problem, presenting a universal interest in all domains where mathematical analysis finds any application, I would indicate the general methods of reasoning which may be usefully applied. The nature of these methods changes according to the topologic or metric character of the properties to be studied. I shall emphasize the necessity for determining carefully this character in order to avoid seeking uselessly by ineffective methods the answers to questions which must be approached from another point of view.

Received March 17, 1939. Translated by Charles and Esther Torrance.

I shall consider only sets and functions represented by points in Cartesian spaces.

Any property of a set is topological if it is preserved under each topologic transformation of the space, that is, under each continuous transformation with a continuous inverse. If the space is linear, such a transformation is continuous and everywhere increasing or everywhere decreasing. If the function f is represented by a point in p-dimensional space, and if the independent variable x is represented by a point in n-dimensional space, then a property of the function $f(x)$ is topological with respect to f or to x if it remains invariant under a topologic transformation of the space of f or x, respectively. The property is doubly topological if it is invariant with respect to both transformations simultaneously.

A property of a set is metrical of the p-th order if it is invariant under each topological transformation of the space which, with its inverse, possesses differential coefficients continuous (or merely finite in some cases) up to at least the p-th order. The properties of a function can be metrical with respect to the independent variable, or with respect to the function, or they can be doubly metrical.

For example, the continuity of a function on a closed set is doubly topological. Also the property of being a unique limit or extreme (greatest, lowest) limit of a sequence of continuous functions, or generally, the class of a function in the sense of Baire, are obviously doubly topological notions. But the property of a function $f(x)$ being of bounded total variation on a straight segment (over which x runs) is topological with respect to the independent variable x, but metrical (not topological) with respect to the function. Absolute continuity is topological neither with respect to the variable nor to the function; this property is doubly metrical.

The fundamental topological notions for sets of points are power, accumulation point, closed set, perfect set, and continuity or discontinuity of a set. Let us observe that in Cartesian spaces, the type of closed set useful for the theory of functions is uniquely the closed bounded set. The properties common to all closed sets, bounded or not, are sheer curiosities. Even though the closed non-compact set is of interest in general analysis, it is likely, on the contrary, that if mathematics had confined itself to the study of functions defined in Cartesian spaces, the notion of the closed non-bounded set would have been dropped. In my own studies of functions of real variables, I have always explicitly stated or understood that all the closed sets considered by me were bounded.

If a set has the topological character of being Borelian, it can be obtained from the class of open sets by successive applications of two operations: (1) addition of a denumerable infinity of sets, and (2) formation of the common part of a denumerable infinity of sets, these two operations being repeated a finite number of times, or transfinitely many times up to an arbitrary rank of class II. These sets were called "measurable B" by Lebesgue who showed their fundamental importance for the theory of the functions of Baire. But Lebesgue later

recognized how this name "measurable" was not apropos, for the character of these sets is merely topological, whereas the properly metrical characteristics of a set involve no topological property. The class of a Borelian set is a topological characteristic. So also is the property of being analytic in the sense of Lusin and Souslin.

Let us pass to the metrical notions. We understand measure in the intrinsic sense of Borel-Lebesgue. For some years now, many people seem to have lost sight of the fact that there is a natural measure, independent of all imaginable conventional measures and to which one must always go back. It is the same as with a function f of Cartesian space; it is possible to transform topologically this space in all manners as many times as one wishes. But in order to reason, one must settle on a determined independent variable, running in its Cartesian space and defined by its ordinary rectangular coördinates. The present generation seems to have forgotten what the geometricians had to overcome in order to determine the ratio of any circumference to its diameter, or to determine the area of a parabolic segment, or more recently, to understand from what point of view an everywhere dense set could have the measure zero. The measure of a set has become all that one wants; however, one person may identify it with a probability, and another may choose it in relation to the data, or even the solution of a problem, in order to prove the truth of some metrical properties of this solution.

For example, consider the trajectories defined on a torus by a differential equation of the first order without singularities. When the Poincaré transformation of the torus into itself is used, it is found that in the transformed torus the trajectories possess the metrical transitivity of G. D. Birkhoff and P. A. Smith; that is, the measure of each measurable family of trajectories is 0 or 1, 1 being the area of the torus. But, as is not the case with true, intrinsical measure, nothing a priori prevents the following situation: if two different equations are given for motion on the same torus, and if two families of trajectories each have the measure 1 in its own measure, depending upon its own equation, it may happen that both families have in common a point set of measure 0 in both measure systems. Obviously, the interest in a metrical transitivity of such a nature is very narrow. The metrical transitivity useful in this theory is the one corresponding to the intrinsic notion of measure. Thanks to this condition, it results that if two families of trajectories on the torus respectively satisfy two differential equations of the type considered, and if each of these families possesses a positive measure, then the set of points on the torus through which there does not pass simultaneously a trajectory of each of these families has the measure zero.

It is therefore understood that just as we deal only with functions and variables represented by points in Cartesian spaces, so do we use only the measure of Borel-Lebesgue or that of Carathéodory, this latter being used for the measure of p-th order in space with $q > p$ dimensions.

The metrical properties of sets are: measurability, the property of having

the measure zero (we shall say, of being *thin*), and the property of being complementary to a set of zero measure, and therefore endowed with a positive measure (we shall say, *thick*). None of these properties is topological. A topological transformation can change a set of measure zero into the complement of such a set and can change a measurable set into a non-measurable set.

As to functions, measurability is metrical with respect to the independent variable, topological with respect to the function. A fundamental result of Lebesgue is that the unique greatest and lowest limits of a sequence of measurable functions are measurable.

For a continuous function the existence of a finite first order derivative is a doubly metrical property of the first order, but not topological. The existence of finite derivatives of the p-th order is a doubly metrical property of the p-th order.

Having recalled these notions, which have been widely and well known for a long time, I would now indicate what fundamental methods of reasoning are used by the theory of functions of real variables, and show how these methods vary according to whether the properties studied are topological or metrical.

One must be thoroughly convinced of the fact that a metrical knowledge of a set or function provides only an extremely unprecise and coarse approximation to the set or function. I understand by metrical knowledge of a set its determination up to a set of zero measure to be added to, or subtracted from, the given set. The metrical knowledge of a function is the knowledge of this function everywhere except at most on a set of measure zero.

A continuous function can have a null derivative everywhere except on a thin set without being constant in any interval. Every measurable function differs from a function of class two only on a set of measure zero. On this latter set are therefore united all characteristic properties of the function. In particular, the class of the function is defined on this set.

While the idea of replacing at each point the value of a function by an approximation is good and often advantageous, yet an approximation of the sets where the function is to be considered is rash and barren. It must be well known that a metrical character of a function or set can be used, and often leads to very interesting consequences, only if the topological nature of the function or set has been previously studied sufficiently deeply. And reciprocally, if to some topological characters presented by a category of functions or sets are added, by way of complement, diverse metrical hypotheses, these latter induce interesting metrical consequences which would vanish along with the topological conditions.

The topological notions to be used in the theory of functions of real variables are essentially those of Cantor and Baire. Baire, like Cantor, never considered the measure of a set. Even if he had been unaware of the existence of this notion, nothing of his known thought or research would have been changed.

Lebesgue capitalized upon the idea of approximation of a function, on the one hand for his definition of an integral, and on the other hand to prove Baire's theorem about functions which are limits of continuous functions. But Lebesgue has never placed himself in the extremely strict, definite, and precise point of view of Baire. Yet without Lebesgue's discipline, many problems, chiefly that of the integration of derivatives, were not to be solved.

The fundamental notion in the topological study of functions is that of the density of a set, either on a continuous or (generally) on a perfect set. Few notions are both more simple and more generally misunderstood, even by the best mathematicians. As if to create an irremediable confusion, people have used the word *density* for a metric notion, i.e., that of the ratio of the measure of a set to the measure of the spatial support containing it. I have always reserved the name *epaisseur* (thickness) for this latter ratio in order to save my reader from a confusion of terms which would forever hide from him the sense and the rôle of topological density. Cantor and Baire perfectly defined this notion which P. du Bois-Reymond had in some degree foreseen. We first define non-density.

A set E lying in the space U_n is called *non-dense* (understood: in U_n) if the derived set E' of E does not possess any interior point. If E' contains a sphere of U_n, E is said to be *dense*. If E' is identical with U_n, E is said to be *everywhere dense*.

Let P denote an arbitrary perfect set in U_n. If ξ represents a regular open set of U_n (ξ is a sphere, a parallelopiped, $\cdots$), the portion of P determined by ξ is, by definition, the part of P interior to ξ together with the accumulation points of this part. In order to determine whether a set E is non-dense, dense, or everywhere dense on P, E is to be limited to its part $E_1 = E \cdot P$ situated in P.

E will be called *non-dense* on P if the derived set E' of E does not contain any portion of P.

E will be called *dense* on P if E' contains any portion of P.

E will be called *everywhere dense* on P if E' is identical with P.

The sum (union) of a denumerable infinity of sets non-dense on P cannot be identical with P. Such a set was called by Baire a set of the first category. I propose to call a *residual* of P a set of P which is the complement of a set of the first category. These notions are still topological. A set of the first category is infinitely rare compared with any residual on the same perfect set P. However, the former set may have measure equal to the measure of P, while the latter has the measure zero. The topological "almost everywhere" is the residual. The metric "almost everywhere" of Lebesgue was called by me "a whole thickness". A denumerable infinity of residuals of P still have in common a residual of P.

The fundamental reasoning in the theory of functions of real variables leads to the following result:

Let ω_n be a sequence of open domains in U_n such that each of these domains contains a point a_i of a perfect set P. If the points a_i are everywhere dense

on P, then the set of points of P belonging to infinitely many of the domains ω_n is a residual of P. If no point of P belongs to infinitely many ω_n, then the points of accumulation of the sequence ω_n necessarily form a closed non-dense set K on P, and any arbitrary portion P_1 of P without points in common with K contains no point of ω_n for all i greater than some number N depending upon P_1.

Let $F(x, t)$ be continuous with respect to x (and often with respect to t) for each value of $t \neq t_0$, and let $f(x)$ be defined at each point x of P as a unique or extreme limit of $F(x, t)$ as t approaches t_0. Uniformity of the progression of $F(x, t)$ toward its limit $f(x)$ would entail consequences of continuity. But if this condition is not satisfied, there exists on P a non-dense set K such that on each portion of P disjoint from K, the conditions of convergence or of semi-convergence of $F(x, t)$ to $f(x)$ are, with some given approximation, uniformly realized. For example, if at each point x of P, $f(x) = \overline{\lim\limits_{t \to t_0}} F(x, t)$ is finite, either f is bounded on P, or there exists a closed, non-dense set K on P such that on each portion P_1 of P without points in common with $K, f(x)$ is bounded by $A(P_1)$ independently of x, and it is even the case that $| F(x, t) | < B(P_1)$ if F is continuous both with respect to x and to t.

Let C denote the circumference of the circle $| x | < 1$. If an analytic function $f(x)$ holomorphic on C tends to 0 when x tends radially towards any point whatever of an arc s of C, then there exists on s an arc s_1 such that in the sector of the circle having O as vertex and s_1 as base, $f(x)$ is bounded. Therefore $f(x) = 0$ on s_1, and it follows that $f(x)$ is identically null.

Let us say that the *index* of a perfect linear set P at any of its points ξ is the smallest number α ($\geqq 1$) such that to each $\alpha' > \alpha$ there corresponds a sequence $\xi + l_0, \xi + l_1, \xi + l_2, \cdots$ situated on P, with $1 < | l_n/l_{n+1} | < \alpha'$.

If a given perfect set P has its index finite at each of its points, then either there are two numbers $\beta(P)$, $\eta(P)$, such that for each ξ of P the sequence $\xi + l_i$ can be formed on P with $\alpha' = \beta(P)$, $| l_0 | > \eta(P)$, or there exists on P a closed non-dense set K such that for each portion P_1 of P without points in common with K, the preceding condition is satisfied for two numbers $\beta(P_1)$, $\eta(P_1)$, independent of ξ while varying in P_1.

If P is a perfect set, and if a continuous function $f(x)$ has at each point a finite upper right derivative, that is, if

$$Q(x, h) = \frac{f(x + h) - f(x)}{h} < A(x) \qquad \text{when } h > 0,$$

then either $Q(x, h) < A(P)$ independently of h and x on P, or there exists on P a closed non-dense set K such that on each portion P_1 of P without points in common with K, $Q(x, h) < A(P_1)$ independently of x on P_1.

Let E be a set in the space U_n, let M be an accumulation point of E, and let us call, with Roger, the derived bundle of E at the point M the set of limits of the half-lines MM' when M' tends to M while moving on E. Let Δ denote

the direction of a straight line, and let $s(M, \Delta, \epsilon, \eta)$ denote the sector of the sphere with center M, axis parallel to Δ, half opening ϵ, and radius η. If at each point of a perfect set P, contained in the derived set E', the direction Δ is not represented in the derived bundle of E, then either $\epsilon(P)$, $\eta(P)$ exist such that all corresponding spherical sectors contain no points of E, or there exists on P a closed set K non-dense on P such that, on each portion P_1 of P having no common point with K, the positive numbers $\epsilon(P_1)$, $\eta(P_1)$ exist.

These examples can be multiplied indefinitely, and all these results are obtained by purely topologic methods.

We shall now discuss metric methods. The thing to be proved is usually that the set of points where a given function $f(x)$ has, or fails to have, a given property, is of measure zero. The principles to be employed are as follows:

1. If a set E over which $f(x) > 0$ has positive measure, there exists a positive number δ such that the set over which $f(x) > \delta$ has positive measure.

For example, if $f(x)$ is continuous, if $\Delta_d(f)$ and $\delta_d(f)$ denote respectively the superior and inferior right derivatives of f, and if the set over which $\Delta_d(f) > \delta_d(f)$ has positive measure, then there exist two numbers α, β $(\alpha < \beta)$, such that the set over which both $\delta_d(f) < \alpha$ and $\beta < \Delta_d(f)$ has positive measure.

2. Each set of positive measure contains a perfect set thick in itself, that is, a set each portion of which has positive measure.

In such a set P, one must prove by topological methods the existence of a portion P_1 where the properties verified on the set P are also verified with some sort of approximate uniformity.

Let E be a set in the space U_n, and let H be the set of points M of E where the derived bundle of E passes through three vertices of a spherical triangle of positive area lying on the sphere with center M and radius 1. Then if H has positive measure, there exist, for any arbitrarily small preassigned positive ϵ, three directions A, B, C, defining a spherical triangle with area superior to a positive number independent of ϵ and such that the set of points of H, having in their derived bundles the three directions A, B, C to within ϵ, has positive measure.

3. On a perfect, thick-on-itself set P, attention may often be advantageously paid to the points where P has the thickness 1.

These general methods are easy to employ and often give very remarkable results. I believe that I am the first to have used them. I have studied the derived numbers of the first order, and also the derived numbers of the second order generalized in the sense of Riemann, belonging to a continuous function of one variable. My analysis was always first topological, then metrical. It led me to the methods of integration which permitted me to calculate the primitive function of a given finite extreme derived number, and also to calculate the second primitive of a second generalized derived number of Riemann. This latter result enabled me to determine the coefficients of a trigonometric series with a given sum.

Later, Menchoff employed the same methods for the study of general conditions for the analyticity of a function of a complex variable. Recently, Roger used these methods to obtain remarkable tangential properties of an arbitrary set in Cartesian space. Less systematical, but always fruitful, applications of these ideas have been made by other authors.

Finally, I shall indicate how, in the theory of Cartesian spaces, metrical hypotheses may be profitably grafted onto given topological conditions.

1. Let E be a set such that any two points of it can be joined by a continuum contained in E. This condition is topological. Let us call an *extreme point* of E each point M of E not belonging to any irreducible continuum joining two points A and B of E distinct from M. It is easy to prove the following result:

If the set of extreme points of E has a positive linear measure, then E has an infinite linear measure.

For example, if a dendrite of Wazewski has a linear finite measure, the set of its extreme points has the measure zero. But the set of extreme points can have a positive n-th measure in U_n while the non-extreme points of the dendrite form a denumerably infinite collection of sets each having a finite measure.

2. Let us consider a plane continuum E. Let $k_\epsilon(E)$ denote an arbitrary continuum contained in E with diameter superior to ϵ. We say that the continuum E has the properties θ_1, θ_2, θ_3 when there exists no infinity of sets $k_\epsilon(E)$ having two by two in common: a null set; a denumerable set; a set non-dense on each of them. These conditions are topological.

Each irreducible continuum joining two points of E is an arc of Jordan. Let Γ be a Jordan arc contained in E and having a positive sense. Let $\eta'(\Gamma)$, $\eta''(\Gamma)$ be respectively the sets of points of Γ accessible on the positive side and on the negative side of Γ from any region of the complementary set of E. With the properties θ_1, θ_2, θ_3 the sets $\eta'(\Gamma)$ and $\eta''(\Gamma)$ are respectively: everywhere dense on Γ; non-denumerable on each arc of Γ; complements of a set non-dense on Γ.

Now let us add a metrical condition which involves θ_3 and which defines the property θ_4': E has a finite linear measure. Under this condition, the sets $\Gamma - \eta'(\Gamma)$ and $\Gamma - \eta''(\Gamma)$ have null measures. Therefore $\eta'(\Gamma)$ and $\eta''(\Gamma)$ have in common a set containing a perfect set in each arc of Γ. It is easy to deduce from this result an example of a continuum containing a circumference Γ and such that the two parts of E respectively interior or exterior to C are each homeomorphic to a set of finite length, while E as a whole is not.

INSTITUT HENRI POINCARÉ.

CONFORMAL MAPPING OF MULTIPLY CONNECTED DOMAINS

By R. Courant

The theory of Plateau's and Douglas' problem furnishes powerful tools for obtaining theorems on conformal mapping. Douglas emphasized (1931) that Riemann's mapping theorem is a consequence of his solution of Plateau's problem; then he treated doubly connected domains and in a recent paper (1939) multiply connected domains.[1] With a different method I gave in a paper on Plateau's problem (1937) a proof of the theorem that every k-fold connected domain can be mapped conformally on a plane domain bounded by k circles.[2] The same method can be applied to the proof of the parallel-slit theorem[3] and, as will be shown in the thesis of Bella Manel, to mapping theorems for various other types of plane normal domains. It is the purpose of the present paper[4] first to give a simplification of the method by utilizing an integral introduced by Riemann in his doctoral thesis, and secondly, to prove a mapping theorem of a different character referring to normal domains which are Riemann surfaces with several sheets.

We consider a Riemann surface on a u, v-plane consisting of the interior of k unit circles which are connected in branch points of total multiplicity $2k - 2$; to this surface we affix $p \geqq 0$ full planes with two branch points each. Thus we define a class of domains B with the boundary b on the plane of $w = u + iv$. Now our theorem is: Each k-fold connected domain G in the x, y-plane with the boundary curves $g_1, g_2, \cdots, g_k$ (which we suppose to be rectifiable[5] Jordan curves) can be mapped conformally on a domain B of our class for any fixed p.

In this mapping the branch points on the full planes and one more branch point may be arbitrarily prescribed and, moreover, on each boundary circle b_ν of B a fixed point may be made to correspond to a fixed point of g_ν.

For the case $p = 0$, the theorem was stated by Riemann, according to oral tradition.[6] As is easily seen, the class of domains B depends on $3k - 6$ essential parameters for $k > 2$ and on one parameter for $k = 2$.

Received March 21, 1939.

[1] Annals of Mathematics, vol. 40(1939), pp. 205–298. This paper contains a complete bibliography of Douglas' previous publications on the Plateau problem.

[2] Annals of Mathematics, vol. 38(1937), p. 679 ff., here quoted as (A).

[3] See M. Shiffman's thesis *The Plateau problem for minimum surface of arbitrary topological structure* to appear in the American Journal of Mathematics.

[4] See also a note by the author in the Proceedings Nat. Acad. Sciences, vol. 24(1938), p. 519 ff. Furthermore, reference is made to a detailed paper by the author on Plateau's and Douglas' problem forthcoming in the Acta Mathematica, here quoted as (B).

[5] This assumption is not essential and can easily be dropped.

[6] See Bieberbach, Sitzungsberichte Berliner Math. Ges., vol. 24(1925), p. 6 ff., where a proof is indicated; and Grunsky, Sitzungsberichte Preuss. Akademie Wiss., 1937, p. 40 ff., where another proof is given.

Without restricting the generality we may assume that G is a finite domain and that g_1 encloses the other boundary curves $g_2, \cdots, g_k$. If a_ν is the positive area enclosed by g_ν and a the area of G, then we have

$$a = a_1 - a_2 - \cdots - a_k .$$

1. **The variational problem.** We call a pair of functions $x(u, v)$ and $y(u, v)$, combined as a vector $\mathfrak{x}(u, v)$, admissible if $\mathfrak{x}$ is continuous in $B + b$, maps the boundary circle b_ν monotonically on g_ν so that the coördination of the fixed boundary points and the sense of description with respect to the domain is preserved, and has piecewise continuous[7] first derivatives in B. (It is not required that $\mathfrak{x}$ map B on the domain G.) Then we consider Riemann's integral,

$$R(\mathfrak{x}) = \tfrac{1}{2} \iint_B [(x_u - y_v)^2 + (x_v + y_u)^2]\, du\, dv, \tag{1}$$

and establish the variational problem R: to find an admissible vector and a domain B of our class for which $R(\mathfrak{x})$ attains its minimum value. If this problem is solved, and if the minimum is shown to be zero, then we have the desired conformal mapping by the analytic function $x + iy = f(u + iv)$; and vice versa, such a conformal mapping is characterized by the equation $R(\mathfrak{x}) = 0$, zero evidently being the minimum value.

By writing

$$\begin{aligned} R(\mathfrak{x}) &= \tfrac{1}{2} \iint_B (\mathfrak{x}_u^2 + \mathfrak{x}_v^2)\, du\, dv - \iint_B (x_u y_v - x_v y_u)\, du\, dv \\ &= D(\mathfrak{x}) - a, \end{aligned}$$

where $D(\mathfrak{x}) = D_B(\mathfrak{x})$ is the "Dirichlet integral" and a is the area included in the total boundary g, we obtain the equivalent problem: to minimize the expression $D(\mathfrak{x}) - a$ by a proper choice of a domain B and an admissible vector $\mathfrak{x}$. Since a depends only on G but not on $\mathfrak{x}$ or on the domain B, we finally have transformed the problem R into the following: to minimize the Dirichlet integral $D(\mathfrak{x})$ by a suitable vector $\mathfrak{x}$ and domain B. This problem is our point of departure.

Because of the Dirichlet principle the solution $\mathfrak{x}$, if it exists, must consist of two harmonic functions:

$$x(u, v) = \Re f(w), \qquad y(u, v) = \Re g(w),$$

where $f(w)$ and $g(w)$ are analytic functions of the complex variable $w = u + iv$ in B and $\Re$ means "real part". However, it is essential in our method not to restrict $\mathfrak{x}$ to harmonic vectors but to retain a greater variability. We have to prove two distinct statements.

[7] This means that in each closed subdomain the continuity may be interrupted at a finite number of points or arcs with continuously turning tangent.

(i) For the solution of our variational problem not only the Laplace equation

$$\Delta\mathfrak{x} = \mathfrak{x}_{uu} + \mathfrak{x}_{vv} = 0 \tag{2}$$

but also

$$\varphi(w) = f'(w)^2 + g'(w)^2 = (\mathfrak{x}_u - i\mathfrak{x}_v)^2 = (\mathfrak{x}_u^2 - \mathfrak{x}_v^2) - 2i\mathfrak{x}_u\mathfrak{x}_v = 0 \tag{3}$$

holds, and this characterizes the mapping of B as conformal mapping on G.

(ii) The solution of the variational problem exists.

2. **The variational condition.** Assuming that the variational problem is solved, we shall express the minimum character of the solution $\mathfrak{x}$ and B first by performing, with B fixed, a suitable variation of the minimizing vector $\mathfrak{x}$; and secondly, by performing variations of the domain B, these latter variations simply consisting in displacements of branch points including dissolution of multiple branch points into simple ones.

We first consider a variation of $\mathfrak{x}$ in a small annular ring A adjacent to one of the boundary circles, e.g. b_1, and so small that A is free from branch points. Introducing concentric polar coördinates r and ϑ, we define in B a function $\lambda(r, \vartheta)$ which is continuous in $B + b$, has there continuous first derivatives λ_r, λ_ϑ, and vanishes in B outside of the ring A. Furthermore, if the point F with $r = 1$ and $\vartheta = \vartheta_0$ on b_1 has a prescribed image on g_1, we require $\lambda(r, \vartheta_0) = 0$. Then for sufficiently small values of the parameter ϵ and polar coördinates r, η the vector

$$\mathfrak{z}(r, \eta) = \mathfrak{x}(r, \vartheta)$$

with

$$\eta = \vartheta + \epsilon\lambda(r, \vartheta)$$

is admissible. Hence we have, if $d = D(\mathfrak{x})$ is the minimum value of the Dirichlet integral,

$$D(\mathfrak{z}) \geqq D(\mathfrak{x}) = d.$$

This leads[8] to the variational equation

$$\lim_{r\to 1} \int_0^{2\pi} \lambda(r, \vartheta) r\mathfrak{x}_r\mathfrak{x}_\vartheta \, d\vartheta = 0 \tag{4}$$

which holds uniformly for all λ with derivatives uniformly bounded in a fixed neighborhood of the boundary circle. Introducing the Poisson kernel

$$K(r, \vartheta; Q) = \frac{r^2 - \rho^2}{2\pi} \frac{1}{r^2 - 2r\rho\cos(\vartheta - \alpha) + \rho^2}$$

for the circle of radius r and for an arbitrary chosen point Q inside the circle with polar coördinates ρ and α, we can define a suitable λ as

$$\lambda = K(r, \vartheta; Q) - K(r, \vartheta_0; Q) \tag{5}$$

[8] See (A), p. 694, or (B).

in a small subring of A adjacent to b_1, and $\lambda = 0$ except in the ring A. Then the condition (4) implies that the analytic function

$$\psi(w) = w^2\varphi(w) \tag{6}$$

is regular in the neighborhood of b_1 except possibly for a pole of the first order at F, and that the values of ψ on b_1 are real.

Proof. The imaginary part of $\psi(w)$ is a harmonic function

$$\Im(\psi(w)) = -2r\mathfrak{x}_r\mathfrak{x}_\vartheta = p \tag{7}$$

which for the argument point Q with the polar coördinates ρ, α shall be written simply $p(Q)$. Within the plane circle with the radius r sufficiently near to 1 we consider the regular harmonic function $\pi_r(Q)$ of the point Q which for the boundary $\rho = r$ coincides with $p(Q)$, so that the difference

$$\delta_r(Q) = \pi_r(Q) - p(Q) \tag{8}$$

has the boundary values zero on the circle $\rho = r$. Now the variational relation (4) for λ chosen as in (5) yields, because of the Poisson formula, for fixed ρ and $r \to 1$ uniformly in Q,

$$\pi_r(Q) - cK_0(Q) \to 0, \tag{9}$$

where $K_0(Q) = K(1, \vartheta_0 ; Q)$. The factor c is the mean value of p on a concentric circle $\rho = r$ and does not depend on r. There may be no branch points for $\rho \geqq \rho_1$; then $|\pi_r(Q)|$ and hence $|\delta_r(Q)|$ is bounded on $\rho = \rho_1$ uniformly with respect to all values of r sufficiently near to 1 since $|K_0(Q)|$ and $|p(Q)|$ are bounded on $\rho = \rho_1$. Now by the principle of reflection the function $\delta_r(Q)$ can be extended to the annular ring $\rho_1 \leqq \rho \leqq r^2/\rho_1$ and is absolutely bounded there. Hence for sufficiently small h the quantity $|\delta_r(Q)|$ is uniformly bounded in the ring $1 - 2h < \rho < 1 + 2h$; therefore in the smaller ring $1 - h \leqq \rho \leqq 1 + h$ the first derivatives of $\delta_r(Q)$ are absolutely bounded by a bound M. Consequently we have $|\delta_r(Q)| < Ms$ if Q is nearer to the boundary $\rho = 1$ than a sufficiently small quantity s. Because of

$$|p(Q) - cK_0(Q)| \leqq |\pi_r(Q) - cK_0(Q)| + |\delta_r(Q)| \tag{10}$$

and (9) we conclude for a small ring adjacent to the unit circle b_1 that $p(Q) - cK_0(Q) = H(Q)$ is a harmonic function having boundary values zero there; hence by the principle of reflection $H(Q)$ is regular in a ring including the circle $\rho = 1$. Since $cK_0(Q)$ is the real part of an analytic function with a pole of first order in F and with vanishing real part everywhere else on the unit circle,[9] our statement is proved.

Secondly, we have to set up the variational conditions referring to the variability of the branch points, of which we may assume none is the point $w = 0$.

[9] Namely, of $c(e^{i\vartheta_0} + w)(e^{i\vartheta_0} - w)^{-1}$ if the fixed point F is the point $e^{i\vartheta_0}$.

These conditions are obtained in the following way. B may be transformed into an admissible domain B' with coördinates u', v' by a transformation

$$u = u' + \epsilon\lambda, \qquad v = v' + \epsilon\mu,$$

where λ, μ are continuous and have continuous bounded first derivatives with respect to u, v or u', v' and ϵ is a small parameter. By defining

$$\mathfrak{z}(u', v') = \mathfrak{x}(u, v),$$

we have

$$D_B(\mathfrak{z}) \geqq D_B(\mathfrak{x}) = d$$

which for $\epsilon \to 0$ yields the desired variational conditions. For this we restrict our transformations to those which affect only a neighborhood N of a branch point $w = w_0$, so that $w' = u' + iv'$ is analytic in w for a neighborhood of w_0 included in N and $w = w'$ outside of N. If $w = w_0$ is a branch point of the order r, we choose in this interior neighborhood

$$w' - w_0 = (w - w_0) + \epsilon(w - w_0)^{\nu/(r+1)} \qquad (\nu = 0, 1, \cdots, r - 1)$$

or

$$w' - w_0 = (w - w_0) + i\epsilon(w - w_0)^{\nu/(r+1)} \qquad (\nu = 0, \cdots, r - 1).$$

This leads to the conditions:[10] If R is a closed curve on B including the branch point P, but no other branch point, we have

$$\int_R (w - w_0)^{\nu/(r+1)} \varphi(w)\, dw = 0 \qquad (\nu = 0, \cdots, r - 1) \tag{11}$$

for

$$\varphi(w) = (\mathfrak{x}_u - i\mathfrak{x}_v)^2 = f'(w)^2 + g'(w)^2.$$

This result can be interpreted as follows: The expansion of $\varphi(w)$ around our branch point has necessarily the form

$$\varphi(w) = \sum_{\mu=0}^{\infty} a_\mu (w - w_0)^{\mu/(r+1)} + \sum_{m=1}^{2r} b_m (w - w_0)^{-m/(r+1)}, \tag{12}$$

as is easily seen from the definition of $\varphi(w)$ after a conformal mapping

$$w^* - w_0 = (w - w_0)^{1/(r+1)}$$

of the neighborhood of P on a simple w^*-plane with a regular expansion of df/dw^* and dg/dw^*. In other words, $\varphi(w)$ appears to have a pole of an order as high as $2r$. But the conditions (11) immediately show that $b_{r+1} = b_{r+2} = \cdots = b_{2r} = 0$, and this means that in a branch point P of order r $\varphi(w)$ has a pole of order not exceeding r. This reduction of the bound for the order of the singularity is the final form for our variational condition.

[10] See Shiffman, loc. cit., or Courant (B).

Now the proof of $\varphi(w) = 0$ follows simply. If $\varphi(w)$ and hence $\psi(w)$ is not identically zero, then ψ must have a finite number N of zeros in B and a finite number M of zeros on the boundary b. According to the result above, the total number P of poles does not exceed the total multiplicity of the variable branch points plus the double multiplicity of the fixed branch points. That means in our general case $P \leqq 2k - 1 + 4p$. Because of the factor w^2, we have $2k + 2p$ zeros at the origin, and, moreover, each point at infinity also represents a double zero of $\psi(w)$ because $f(w)$ and $g(w)$ are regular there. Thus we have $N \geqq 2k + 4p$. But $2\pi(N - P)$ is equal to the total increase of the amplitude of $\psi(w)$ if we describe the boundary b in the positive sense, avoiding zeros on b and possibly poles by small half circles in B. Since ψ is real on b, only these half circles contribute to the total variation of the amplitude, each zero the amount $-\pi$, and each pole the amount $+\pi$. If no points on the boundary were fixed, there would be no pole there, and hence $N - P < 0$. Therefore,

$$2k + 4p \leqq P \leqq 2k - 1 + 4p, \tag{13}$$

and this is a contradiction. Hence $\psi(w) = 0$ is proved.

If there is a point F fixed, e.g., on the circle b_1, the function $\psi(w)$ has a pole of, at most, the first order there, contributing $+\pi$. But since ψ is real on b_1 and continuous except at F, where it must approach $+\infty$ and $-\infty$ respectively from the two sides, there must necessarily be at least one zero of $\psi(w)$ on b_1, contributing $-\pi$ so that again $N - P \leqq 0$. Therefore our conclusion subsists.

It is apparent that this method for establishing mapping theorems permits many other applications which, however, are omitted here. We mention only the case in which the unit circles are replaced by concentric circles with non-determined radii. In this case more branch points can be fixed.

3. **Solution of the variational problem.** To solve the variational problem we consider a minimizing sequence of domains B_n and corresponding vectors $\mathfrak{x}_n$ so that $D_{B_n}(\mathfrak{x}_n) \to d$, where d is the lower limit of the Dirichlet integral. We say that a sequence $\mathfrak{x}_n$ tends to separation if there is a closed continuous curve $\beta = \beta_n$ in B_n, not capable of contraction to a point there, so that the image $\gamma = \gamma_n$ of β in the x, y-plane has a length $\epsilon = \epsilon_n$ tending to zero. We may assume without restricting the generality that β separates B into the domains B_n' and B_n'' of which b_1 and b_2 are boundary curves, respectively. If there exists a minimizing sequence not tending to separation, the variational problem is solvable. For the simple proof of this theorem we refer to the paper (A) or (B). Therefore what remains to be shown is that for our minimizing sequence a fixed positive lower bound for ϵ can be found. This is the crucial point, of which we shall dispose on the basis of a general continuity method by utilizing the Riemann integral $R(\mathfrak{x})$.

We first derive an intuitively plausible inequality. Suppose we have in $B = B_n$ with the admissible vector $\mathfrak{x}$ a closed simple curve β as described above.

The image of B by $\mathfrak{x}$ may cover parts of the plane in G or outside of G several times. If ϵ is the length of the image γ of β, then we have for the "area"

$$\eta = \frac{1}{2}\int_B x\,dy - y\,dx$$

enclosed by γ the inequality

$$|\eta| < \epsilon^2. \tag{14}$$

Now we consider the integral

$$R'(\mathfrak{x}) = \frac{1}{2}\iint_{B'} [(x_u - y_v)^2 + (x_v + y_u)^2]\,du\,dv = D'(\mathfrak{x}) - (a_1 - \cdots - \eta),$$

where $D'(\mathfrak{x})$ refers to the domain B'; in the last expression on the right side the area a_2 is missing. On account of $R'(\mathfrak{x}) \geqq 0$ we have therefore by $D(\mathfrak{x}) \geqq D'(\mathfrak{x})$ the inequality

$$D(\mathfrak{x}) \geqq a_1 - \cdots - \eta \geqq a + a_2 - \eta,$$

where a is the area of G. Hence

$$D(\mathfrak{x}) \geqq a + a_2 - \epsilon^2.$$

Since $\mathfrak{x}$ is a member of a minimizing sequence, we may for an arbitrarily given positive number q assume that

$$D(\mathfrak{x}) < d + q,$$

where d is the lower limit of the variational integral D. Then we have

$$d + q > a + a_2 - \epsilon^2. \tag{15}$$

This is the basic inequality.

Our continuity method starts with the continuity theorem.[11] We consider a σ-deformation of the x, y-plane

$$x' = x + \xi(x, y), \qquad y' = y + \eta(x, y),$$

where ξ, η and their first derivatives are continuous in the plane and have absolute values less than σ. Then each vector $\mathfrak{x}(u, v)$ in the u, v-domain is transformed into another vector $\mathfrak{x}'(u, v)$ in the same domain having the components x', y' and mapping the boundary g of B onto the boundary of g' of a slightly deformed domain G' if σ is sufficiently small. The lower limit of the variational integral $D(\mathfrak{x}')$ may be d' if in the conditions G is replaced by G'. Then the theorem states the inequality

$$d' \leqq d(1 + 2\sigma)^2$$

between the two lower limits d and d'.

[11] See (A), pp. 695 and 707 ff.

The proof is immediate. We have

$$D(\mathfrak{x} - \mathfrak{x}') = \tfrac{1}{2} \iint_B (\xi_u^2 + \xi_v^2 + \eta_u^2 + \eta_v^2)\, du\, dv$$

$$= \tfrac{1}{2} \iint_B [(\xi_x x_u + \xi_y y_u)^2 + \cdots]\, du\, dv.$$

Hence by Schwarz' inequality

$$D(\mathfrak{x} - \mathfrak{x}') < 2\sigma^2 \iint_B (x_u^2 + x_v^2 + y_u^2 + y_v^2)\, du\, dv = 4\sigma^2 D(\mathfrak{x})$$

and by

$$(\sqrt{D(\mathfrak{x})} - \sqrt{D(\mathfrak{x}')})^2 \leqq D(\mathfrak{x} - \mathfrak{x}'),$$

we have finally

$$\sqrt{D(\mathfrak{x}')} < (1 + 2\sigma) \sqrt{D(\mathfrak{x})}.$$

Our statement follows from this because $D(\mathfrak{x})$ may be chosen arbitrarily near to d and $D(\mathfrak{x}')$ is certainly not less than d'.

As a corollary we observe that the conditions in the theorem can be eased by admitting continuous functions $\xi(x, y)$, $\eta(x, y)$ for which the continuity of the first derivatives may be interrupted along a finite number of straight lines. For, we may assume $\mathfrak{x}(u, v)$ as a harmonic vector; then to such a straight line $ax + by + c = 0$ there corresponds in the u, v-plane an equipotential line of the harmonic function $ax(u, v) + by(u, v)$. Hence discontinuities of the derivatives of $\mathfrak{x}'(u, v)$ are of an admissible character, occurring only on such analytic lines.

We furthermore make the preliminary remark that the lower limit d is equal to the area of G, provided that the problem is solved, as follows immediately by §2. Conversely, if B is mapped conformally on G, then the mapping vector $\mathfrak{x}$ solves the variational problem. For we always have $D(\mathfrak{x}) \geqq a$, while in this case $D(\mathfrak{x}) = a$, $R(\mathfrak{x}) = 0$.

A third preliminary remark is: there exists a specific domain G_0 for which the problem certainly can be solved. For we take the special domain $B = B_0$ with two branch points, of multiplicity $k - 1 + p$ each, at $w = \frac{1}{2}$ and $w = -\frac{1}{2}$. Then since

$$r + is = \left(\frac{2w + 1}{2w - 1}\right)^{1/(k+p)},$$

B is mapped on a domain in an r, s-plane which by a further linear transformation can be transformed into a domain G_0 of our type with analytic boundaries.

It is sufficient to restrict ourselves to domains G with piecewise smooth[12]

[12] These can be defined as follows. The boundary may consist of a finite number of arcs each of which, after a suitable rotation of the coördinate system, can be expressed in the form $y = g(x)$, where $g'(x)$ is continuous in the corresponding closed x-interval.

boundaries, e.g. polygons, because for the most general Jordan boundaries the result then follows by a simple passage to the limit.

Now we connect G_0 with G by a sequence of k-fold connected domains G_1, $G_2, \cdots, G_L = G$, each of which is transformed into the next by a σ-transformation. By an elementary construction this can immediately be effected with any σ by means of transformations where discontinuities of ξ_x, ξ_y, η_x, η_y only occur on straight lines. The boundaries g^m of G_m consist of k curves g_ν^m, and the corresponding areas may be denoted by a^m and a_ν^m, respectively. For all m and ν and arbitrary σ we may assume that

$$a_\nu^m \geqq 4q,$$

$4q$ being a fixed positive quantity, namely, the area of a circle small enough to fit in the interior of $g_1, \cdots, g_k$ and all the g_ν^m. The lower limit of our variational problem concerning G^m is denoted by d_m.

For the first element G_0 of the chain the problem is solved; therefore we have

$$d_0 = a^0.$$

We prove—and this is the main point of the reasoning—that solvability for G_m implies solvability for G_{m+1}.

For this we choose σ so small that

$$(1 + 2\sigma)^2 - 1 < q \tag{16}$$

and

$$a^{m+1} > a^m - q. \tag{17}$$

Furthermore, by a similarity transformation of the plane, we may assume for convenience that

$$a^m < 1 \qquad \text{for all } m. \tag{18}$$

Solvability for G_m implies

$$d_m = a^m = a_1^m - a_2^m - \cdots - a_k^m. \tag{19}$$

The continuity theorem yields because of (16), (17), (18)

$$d_{m+1} \leqq a^m(1 + q) \leqq a^m + q \leqq a^{m+1} + 2q,$$

therefore

$$d_{m+1} - a^{m+1} \leqq 2q. \tag{20}$$

Now the inequality (15) is applied for the domain G_{m+1}. We obtain

$$d_{m+1} + q > a^{m+1} + a_2^{m+1} - \epsilon^2,$$

and hence by $a_2^{m+1} > 4q$ we have

$$d_{m+1} > 3q + a^{m+1} - \epsilon^2$$

or

$$\epsilon^2 > 3q + a^{m+1} - d_{m+1} .$$

By (20)

$$\epsilon^2 > q$$

results, and this gives the desired lower bound for ϵ. Thus the solvability of the problem for G_{m+1} is assured.

Starting with G_0 we therefore find that the problem for $G = G_L$ is solvable. This completes the proof of our mapping theorem.

NEW YORK UNIVERSITY.

DUALITY AND COMMUTATIVITY OF GROUPS

By Reinhold Baer

Two groups are duals of each other if there exists an anti-isomorphism between their lattices of subgroups. The existence of a dual implies that all the elements are of finite order. Not every group possesses a dual since there do not exist duals of Hamiltonian groups. To be a dual and to be Abelian are equivalent properties of finite groups generated by elements of prime order p; and this statement is but one in a larger class of theorems connecting duality and commutativity.

Abelian groups possess duals if, and only if, they are self-dual; and a necessary and sufficient condition for self-duality of an Abelian group is the absence of elements of infinite order together with the finiteness of its primary components. Thus the class of self-dual Abelian groups proves to be exactly the same as the class of those Abelian groups—determined in an earlier note[1]—which admit an operation mapping the subgroups upon isomorphic quotient-groups.

1. A *dualism* between the groups G and H is a one-one correspondence $\mathbf{d}$, mapping the set of all the subgroups of G upon the whole set of subgroups of H in such a way that

$$S \leqq T \text{ if, and only if, } T^{\mathbf{d}} \leqq S^{\mathbf{d}}.$$

Such a dualism maps the cross-cut (join-group) of a set of subgroups upon the join-group (cross-cut) of the set of the corresponding subgroups; and it maps in particular G upon the identity in H and the identity in G upon H. Two groups are *duals* of each other if there exists a dualism between them.

The inverse operation of a dualism is again a dualism between the same groups, whereas the product of two dualisms (if it exists) is a so-called *subgroup-isomorphism*,[2] as it maps a subgroup of a subgroup upon a subgroup of the corresponding subgroup.

If $\mathbf{d}$ is a dualism between the groups G and H, and if in particular $G = H$, then $\mathbf{d}$ is an *autodualism* of G and the group G is *self-dual.*

A dualism $\mathbf{d}$ between the groups G and H induces dualisms in all the quotient-groups of H and in all those subgroups of G which it maps upon normal subgroups of H. This principle will be used very often.

The following theorem is known[3] and may be stated for future reference.

Received March 17, 1939; presented to the American Mathematical Society, April 7-8, 1939.

[1] Reinhold Baer, *Dualism in Abelian groups*, Bull. Amer. Math. Soc., vol. 43(1937), pp. 121–124; cited as Dualism.

[2] Cf. R. Baer, Bull. Amer. Math. Soc., vol. 44(1938), pp. 817–820; Amer. Jour. of Math., vol. 61(1939), pp. 1–44; cited as Isomorphism.

[3] See Dualism.

THEOREM E. *If G is an Abelian group without elements of infinite order whose primary components*[4] *are finite, there exists an autodualism* $\mathbf{d}$ *of G with the following properties*:

(a) *If S is a subgroup of G, then S and $G/S^{\mathbf{d}}$ as well as $S^{\mathbf{d}}$ and G/S are pairs of isomorphic groups.*

(b) *There exists an automorphism of G, mapping S upon T, if, and only if, there exists an automorphism of G, mapping $S^{\mathbf{d}}$ upon $T^{\mathbf{d}}$.*

It is our object to prove the following theorems.

THEOREM A. *If G is an Abelian group, then the following two conditions are necessary and sufficient for the existence of a dualism between G and some other group (which may be Abelian or not, equal to G or not)*:

(1) *G does not contain elements of infinite order*;

(2) *each primary component of G is finite.*

THEOREM B. *There exists a dualism between the Abelian group G and the group H if, and only if,*

(i) *both G and H are self-dual;*

(ii) *G and H are subgroup-isomorphic.*

It should be noted that it may happen that a non-commutative group is a dual of an Abelian group. For it is a consequence of Theorem E that every finite Abelian group F is self-dual; and it is known that there exist finite Abelian groups F which are subgroup-isomorphic with non-commutative groups.[5]

The following special case of Theorem B seems to be noteworthy:

An Abelian group possesses a dual if, and only if, it is self-dual.

The author has not been able to decide whether the hypothesis that the group be Abelian is necessary for the validity of this statement. The existence of a dual, however, causes the structure of the group to be "almost Abelian", as may be seen from the theorems proved in the §§5–7, the most notable one of these being the following

THEOREM C. *A finite p-group G is Abelian if, and only if,*

(a) *there exists a dual of G;*

(b) *non-Abelian quotient-groups S/T of subgroups S of G are not subgroup-isomorphic with Abelian groups.*

2. In this section we prove some necessary conditions for the existence of duals.

THEOREM 2.1. *If a group possesses a dual, then all its elements are of finite order.*

Proof. Suppose that $\mathbf{d}$ is a dualism between the groups G and H (which may be commutative or not), and that Z is an infinite cyclic subgroup of G. The subgroups Z^{2^i} for $0 \leq i$ form a properly descending chain of subgroups of Z (and therefore of G) whose cross-cut is 1. The subgroups $W_i = (Z^{2^i})^{\mathbf{d}}$ form

[4] A primary component of a group consists of all those elements whose orders are powers of the fixed prime number p.

[5] See Isomorphism.

consequently a properly ascending chain of subgroups of the group H whose set-theoretical join is H, since H is certainly the smallest subgroup of H which contains all the W_i, and since the set of all the elements which occur in at least one W_i is a subgroup of H. If $V = (Z^3)^{\mathbf{d}}$, then $W_0 < V$ and there does not exist any subgroup different from V and W_0 between V and W_0, since there does not exist any subgroup different from Z and Z^3 between Z and Z^3. In particular, there exists an element v in V which is not contained in W_0 and which is contained in at least one subgroup W_k, since H is the join of these W_k. If v is an element in W_k, then $V \leqq W_k$, since $W_0 \leqq W_k$ and since V is the subgroup of H, generated by v and the elements in W_0. Applying the dualism $\mathbf{d}^{-1}$ upon this inequality, it follows that $Z^{2^k} \leqq Z^3$; and this is impossible.

LEMMA 2.2. *If there exists a dual of the group G, and if there exists a subgroup $K \neq 1$ of G which is contained in every proper subgroup of G, then G is a finite cyclic group.*

Proof. If $\mathbf{d}$ is a dualism between G and the group H, then every subgroup $S \neq H$ of H is contained in $K' = K^{\mathbf{d}}$ and K' is different from H, since $K \neq 1$. Hence there exists an element w in H which is not contained in K'. The cyclic subgroup of H which is generated by w is not a subgroup of K' and is consequently equal to H so that H is a cyclic group. Since G is a dual of H, it follows from Theorem 2.1 that H is a finite cyclic group. H is therefore by Theorem E self-dual. This implies that G and H are not only duals, but are even subgroup-isomorphic. But it is known[6] that only finite cyclic groups are subgroup-isomorphic with finite cyclic groups; and this proves that G is a finite cyclic group.

COROLLARY 2.3. *Groups of type p^∞ do not possess duals.*[7]

This is a simple consequence of Lemma 2.2, since groups of type p^∞ are not cyclic, but contain a subgroup of order p which is contained in every subgroup different from 1.

COROLLARY 2.4. *Hamiltonian groups do not possess duals.*

Proof. If a group possesses a dual, then so does each of its quotient-groups. Every Hamiltonian group is a direct product of a quaternion-group and some Abelian group so that every Hamiltonian group is homomorphic to a quaternion-group. Thus the existence of duals to Hamiltonian groups would imply the existence of duals to quaternion-groups. But it is a consequence of Lemma 2.2 that a quaternion-group does not possess a dual, since such a group is not cyclic but contains a subgroup of order 2—namely, its central—which is a subgroup of every proper subgroup.

Remark. In order to delineate the scope of applicability of Lemma 2.2 one

[6] Isomorphism, Theorem 2.4, p. 6.

[7] A group of type p^∞ is an Abelian group, generated by elements a_i, so that a_1 is of order p and $a_i^p = a_{i-1}$.

has to determine all the groups which contain (exactly) one smallest proper subgroup, i.e., a proper subgroup which is contained in every proper subgroup. This may be done as follows.

This class of groups certainly contains the cyclic groups $Z(p^n)$ of order a power of a prime p^n for $1 < n$, and the groups $Z(p^\infty)$ of type p^∞. There are furthermore the dihedral groups $D(n)$ for $0 < n$ which are generated by two elements u and v, subject to the relations

$$u^{2^{n+1}} = 1, \qquad u^{2^n} = v^2, \qquad vuv^{-1} = u^{-1};$$

and the *infinite dihedral group* $D(\infty)$ which is generated by elements $u_1, u_2, \cdots, u_i, \cdots, v$, subject to the relations

$$u_i^2 = u_{i-1}, \qquad u_1^2 = v^2, \qquad u_1^4 = 1, \qquad vu_iv^{-1} = u_i^{-1};$$

$D(\infty)$ is clearly the join of the groups $D(n)$, generated by v and u_n; and the element v^2 generates the only subgroup of order 2.

Suppose now that the group G contains exactly one smallest proper subgroup. If G is finite, then it is a well-known theorem[8] that G is either a group $Z(p^n)$ or a group $D(n)$. If G is infinite, then let us assume that any finite number of elements in G generates a finite subgroup of G. (Whether or not this hypothesis is really needed the author has not been able to decide.) If the prime number p is the order of the uniquely determined smallest proper subgroup of G, then it follows from the facts mentioned so far that G contains one and only one cyclic subgroup of order p^n for $2 < n$. If either p is odd or G is Abelian, then G is the join of these cyclic subgroups, as all the elements in G are of order a power of p, i.e., G is a group $Z(p^\infty)$. If $p = 2$, but G is not Abelian, the above argument shows that G contains a uniquely determined subgroup $Z(2^\infty)$ and is the join of an ascending chain of subgroups $D(n)$, i.e., G is a group $D(\infty)$.

It is perhaps of interest to note that the following property of groups is equivalent to the one which we investigated just now.

There exists a proper subgroup of the group G which is a subgroup of every proper subgroup if, and only if, there exists a proper subgroup C of G with the property:

(*) *If S is a subgroup of G, then either $S \leqq C$ or $C \leqq S$.*

Since a smallest subgroup (in the strong sense) is a subgroup, contained in every proper subgroup, the smallest subgroups satisfy (*). Hence we need only show that the existence of a group with the property (*) is sufficient for the existence of a smallest subgroup. Suppose therefore that C is a proper subgroup of G and that C satisfies (*). There exist elements in G which are not contained in C, since $C < G$. If u is any element in G which is not contained in C, and if U is the subgroup generated by u, then it follows from (*) that $C < U$, since $U \nleqq C$. There exists therefore an integer $n > 1$ so that $C = U^n$. Suppose

[8] W. S. Burnside, *Theory of Groups of Finite Order*, 2d ed., 1911, Theorems V and VI on pp. 131–132.

now that u is of infinite order. Then let p be any number which is relatively prime to n. The subgroup U^p of U would then neither contain $U^n = C$ nor would U^p be a subgroup of $U^n = C$. As this is impossible, u is of finite order. If the order of u is the product hk of two numbers h and k which are relatively prime, then it is impossible that both u^h and u^k are in C, since otherwise u would be in C. If, for example, u^h is not in C, then $C \leqq U^h$ and U^h is of order k. Since h and k are relatively prime, it follows that 1 is the cross-cut of U^h and U^k; and this implies that 1 is the cross-cut of U^k and C. Since $C \neq 1$, it follows from the property (*) that $U^k = 1$ so that u is of order k, and $h = 1$. Now it follows that all the elements in G are of order a power of a prime number. As in particular C is a cyclic group of order a power of a prime number, C contains a subgroup D of order a prime number. If S is a proper subgroup of G, then either $D \leqq C \leqq S$ or $S < C$. In the latter case, however, $D \leqq S$, since D is a subgroup of every proper subgroup of C. This completes the proof.

THEOREM 2.5. *There does not exist a dual of an infinite Abelian group all of whose elements $\neq 1$ have the same finite order.*

Proof. If G is an Abelian group all of whose elements $\neq 1$ have the same finite order, then G is a direct product of cyclic groups of order p, where p is a prime number. If G is infinite, the number of factors in such a decomposition is the same as the number of elements in G and the same as the number of cyclic subgroups. This number may be denoted by $r(G)$. If $r(G)$ is infinite, one verifies easily that the number of subgroups of index p in G is exactly $2^{r(G)}$ (e.g., by considering the homomorphic mappings of G upon a cyclic group of order p).

Suppose now that $\mathbf{d}$ is a dualism between G and the group H (which may be commutative or not). Let u and v be two elements in H and suppose that the subgroup W of H which is generated by u and v is not cyclic. If U is the subgroup generated by u, and V the subgroup generated by v, then U and V are proper subgroups of W, and W is the smallest subgroup containing both U and V. This implies that $W' = W^{\mathbf{d}^{-1}}$ is the cross-cut of $U' = U^{\mathbf{d}^{-1}}$ and $V' = V^{\mathbf{d}^{-1}}$.

Since $\mathbf{d}$ induces a dualism between G/U' and U, it follows from Theorem 2.1 that U is a finite cyclic group. Hence it follows from Theorem E that U is self-dual; and G/U' and U are therefore subgroup-isomorphic. Since the subgroup-isomorphic image of a finite cyclic group is itself a finite cyclic group,[9] G/U' is a finite cyclic group. But cyclic quotient-groups of G are all of order p so that there is no proper subgroup between U' and G. Hence U does not possess any proper subgroup and this shows that *all the elements $\neq 1$ in H are of order a prime number.*

This implies, in particular, that U and V have only 1 in common, and G is consequently the smallest subgroup of G containing both U' and V'. Hence G/W' is a direct product of two cyclic groups of order p. It follows from Theorem E that there exists an autodualism of G/W'. Since furthermore $\mathbf{d}$

[9] Cf. footnote 6.

induces a dualism between G/W' and W, this proves the existence of a subgroup-isomorphism between G/W' and W. Applying now a general theorem[10] on subgroup-isomorphisms of direct products of groups of order p, we find that

W is either a direct product of two cyclic groups of order p; or

W contains a normal subgroup of order p, and all the other proper subgroups of W are of order a prime number q which divides $p - 1$.

It is a consequence of this result that the elements of order p form (together with 1) a subgroup P of H which is a direct product of cyclic groups of order p and which is either equal to H or of index a prime number q in H.

$\mathbf{d}$ induces a dualism between $P' = G/P^{\mathbf{d}^{-1}}$ and P. As P is infinite, P' is infinite. Thus the number of subgroups of order p in P' is $r(P')$, and the number of subgroups of index p in P' is $2^{r(P')}$. This is clearly the number of subgroups of order p in P and the number of subgroups of index p in P is therefore $2^{2^{r(P')}}$. But it follows from the existence of $\mathbf{d}$ that the number of subgroups of index p in P is the same as the number of subgroups of order p in P'. Hence $r(P') = 2^{2^{r(P')}}$ and this is impossible. Consequently G cannot be infinite, and the proof is complete.

3. The proof of the Theorems A and B is now fairly simple. Suppose first that G is an Abelian group, and that $\mathbf{d}$ is a dualism between G and the group H which may be commutative or not.—It may be noted that our arguments would be much simpler if we assumed that H is Abelian.—Then it is a consequence of Theorem 2.1 that G does not contain elements of infinite order. G is therefore the direct product of its primary components G_p, where G_p consists of all the elements in G whose order is a power of the prime number p. If Z is a subgroup of type p^∞ of G, then Z is a direct factor[11] of G, i.e., $G = Z \times Z'$ and G/Z' is a group of type p^∞. But $\mathbf{d}$ induces a dualism between G/Z' and $Z'^{\mathbf{d}}$ and this contradicts Corollary 2.3. Hence G does not contain a subgroup of type p^∞. If p is any prime number, then G_p/G_p^p and G/G^p are essentially the same groups. Again $\mathbf{d}$ induces a dualism between G/G^p and $(G^p)^{\mathbf{d}}$; and as the elements $\neq 1$ in G/G^p are of order p, it follows from Theorem 2.5 that G/G^p is finite. Hence G_p does not contain subgroups of type p^∞, and G_p/G_p^p is finite; and therefore it follows from general theorems on primary Abelian groups[12] that G_p is finite. Thus it has been proved that the conditions in Theorem A are necessary; and that these conditions are sufficient follows from Theorem E.

That the conditions of Theorem B are sufficient is an obvious consequence of the fact that the product of an autodualism and a subgroup-isomorphism is a dualism. If on the other hand the Abelian group G and the group H are duals of each other, it follows from Theorem A that G satisfies the conditions of Theorem E so that G is by Theorem E self-dual. G and H are therefore subgroup-isomorphic, since the product of an autodualism and of a dualism is a

[10] Isomorphism, Theorem 8.1, p. 14 and Theorem 11.2, pp. 23–24.

[11] R. Baer, Annals of Math., vol. 37(1936), (1; 1), p. 766.

[12] Cf., e.g., R. Baer, Amer. Jour. of Math., vol. 59(1937), pp. 99–101.

subgroup-isomorphism; and finally H is self-dual, since H is subgroup-isomorphic with a self-dual group.

4. We say that the subgroup S of G is *chained to* G, if the subgroups between S and G form a finite chain, i.e., if the subgroups U and V of G both contain S, then either $U \leqq V$ or $V < U$; and the number of subgroups between S and G is finite. If T is any subgroup of the group G, then there are always subgroups S between T and G which are chained to G, e.g., G itself. We denote by $B(T)$ the cross-cut of all those subgroups between T and G which are chained to G.

LEMMA 4.1. (a) *Suppose that* $\mathbf{d}$ *is a dualism between the groups* G *and* H. *The subgroup* S *of* G *is chained to* G *if, and only if,* $S^{\mathbf{d}}$ *is a cyclic group of order a power of a prime number.*

(b) *If* T *is a subgroup of the group* G, *and if there exists a dual of* G, *then* $T = B(T)$.

Proof. To prove (a) one has only to remark that S is chained to G if, and only if, the subgroups of $S^{\mathbf{d}}$ form a finite chain; and that the cyclic groups of order a power of a prime number are characterized by this last property.[13] In order to prove (b) let $\mathbf{d}$ be a dualism between G and the group H. Then it is a consequence of Theorem 2.1 that there are no elements of infinite order in H. This implies in particular that every subgroup of H is generated by its finite cyclic subgroups. Since every finite cyclic group is generated by cyclic groups of order a power of a prime number, it follows that every subgroup of H is generated by cyclic groups whose order is a power of a prime number. It is a consequence of (a) that the subgroup $B(T)^{\mathbf{d}}$ of $T^{\mathbf{d}}$ is generated by all those cyclic subgroups of $T^{\mathbf{d}}$ whose order is a power of a prime number. Hence $B(T)^{\mathbf{d}} = T^{\mathbf{d}}$ and $B(T) = T$.

The most important special case of (b), as far as our applications are concerned, is $T = 1$. It may be stated separately:

If there exists a dual of the group G, *there exists corresponding to every subgroup* $A \neq 1$ *of* G *a subgroup* B *of* G *which is chained to* G *and does not contain* A.

5. The theorems in this section will be needed in the proof of Theorem C. But they are not special cases of this theorem.

THEOREM 5.1. *If* G *is a group with Abelian central quotient-group, if* G *is generated by elements of order a prime number* p, *and if there exists a dual of* G, *then* G *is a finite Abelian group.*

Proof. Let us denote by Z the central and by C the commutator subgroup of G. Then $C \leqq Z \leqq G$. As both Z and G/Z are Abelian groups which are generated by elements of order p, all their elements $\neq 1$ are of order p;[14] and they are direct products of cyclic groups of order p.

[13] Isomorphism, (2.1).

[14] See, e.g., R. Baer, Trans. Amer. Math. Soc., vol. 44(1938), pp. 357–386.

The proof will now be conducted in two steps, the first of them dealing with a special case of the general theorem whereas the second step contains the reduction of the general statement to the special case.

I. Let us assume that $C = Z$ is a cyclic group of order p; and that there exists a dualism $\mathbf{d}$ between G and the group H. $C \neq 1$ implies that G is not Abelian so that $Z \neq G$.

There exists by Lemma 4.1 a subgroup U of G which is chained to G and which does not contain Z. Since $Z = C$ is of order p, U and C have but the identity in common so that in particular U is Abelian and so that the subgroup U', generated by U and Z, is the direct product of U and Z and is therefore Abelian too. Since U is chained to G, and since U' is between U and G, it follows that U' is chained to G. Since U' is Abelian and G is not, $U' < G$. Since $C \leqq U'$, G/U' is Abelian and therefore cyclic. Since the cyclic subgroups and quotient-groups of G/Z are all of orders 1 or p, it follows that G/U' is of order p.

Let v be any element of G which generates G mod U'. Then there exists an element u in U so that $c = uvu^{-1}v^{-1}$ is different from 1, since the elements in $Z = C$ are permutable with v. Consequently c generates C. If $x \neq 1$ is an element in U, then x is not in Z. Since U' is Abelian, and since v generates G mod U', it follows that $vxv^{-1}x^{-1} = c^i$, where i is relatively prime to p. Now it follows from general properties of groups with Abelian central quotient-group[14] that

$$v(xu^{-i})v^{-1}(xu^{-i})^{-1} = vxv^{-1}x^{-1}(vuv^{-1}u^{-1})^{-i} = 1,$$

so that xu^{-i} as an element in U is permutable with the elements in U' as well as with v; i.e., xu^{-i} belongs to the central of G. But the cross-cut of Z and U is 1 so that $xu^{-i} = 1$ or $x = u^i$. U is consequently a cyclic group of order p and G/Z is a direct product of two cyclic groups of order p. Since the group G is generated by elements of order p, it is possible to choose the element v so that its order is p. Thus it has been shown that

G is a group of order p^3 which is generated by two elements u and v subject to the relations:

$$1 = u^p = v^p = (uvu^{-1}v^{-1})^p;$$

$$uuvu^{-1}v^{-1} = uvu^{-1}v^{-1}u, \qquad vuvu^{-1}v^{-1} = uvu^{-1}v^{-1}v.$$

If $p \neq 2$, then all the elements $\neq 1$ in G are of order p; if $p = 2$, then u, v and $(uv)^2$ are the elements of order 2 and $(uv)^{\pm 1}$ are the elements of order 4.

A subgroup M of G may be called a greatest subgroup of G, if $M < G$, and if there does not exist a subgroup N so that $M < N < G$. That subgroups of index p in G are greatest subgroups in G is clear. If the index of a subgroup S in G is p^2, then either $S = Z$ or S is of order p and the cross-cut of Z and S is 1. Z itself is not even chained to G, since G/Z is a direct product of two cyclic groups of order p. If $S \neq Z$, then the subgroup S', generated by S and Z, is their direct product. Hence S' is of index p in G. Thus it has been shown that *the greatest subgroups of G are exactly the subgroups of index p in G. The*

subgroups of index p^2 in G which are different from Z are chained to G, but are not greatest subgroups of G.

Since Z is contained in every subgroup of index p in G, and since G/Z is a direct product of two cyclic groups of order p, it follows that *Z is the cross-cut of all the greatest subgroups of G.*

Since G is not cyclic, it follows that $Z^{\mathbf{d}}$ is generated by all the cyclic subgroups of H whose order is a prime number and the order of every element $\neq 1$ in $Z^{\mathbf{d}}$ is a prime number. This implies, in particular, that $Z^{\mathbf{d}}$ is a normal subgroup of H; and $H/Z^{\mathbf{d}}$ is a cyclic group of order a prime number, since Z is of order p. If U is a subgroup of order p of G, and if $U \neq Z$, then $U^{\mathbf{d}}$ is a cyclic group of order a square of a prime number. Since $U^{\mathbf{d}}$ is not part of $Z^{\mathbf{d}}$, $U^{\mathbf{d}}$ generates H mod $Z^{\mathbf{d}}$. If q is the order of $H/Z^{\mathbf{d}}$, this implies that $U^{\mathbf{d}}$ is of order q^2.

$Z^{\mathbf{d}}$ contains exactly $p + 1$ subgroups of order a prime number, since G/Z contains exactly $p + 1$ subgroups of order p. If $p = 2$, this implies—together with the fact that all the proper subgroups of $Z^{\mathbf{d}}$ are of order a prime number—that $Z^{\mathbf{d}}$ is a direct product of two cyclic groups of order 2 and that $Z^{\mathbf{d}}$ is therefore Abelian.[15] If $p \neq 2$, then every greatest subgroup M of G contains a subgroup U such that $U \nleq Z$ and such that U is of order p. Such a subgroup U of G is chained to G, is of index p^2 in G and is therefore not a greatest subgroup of G. Consequently, every $M^{\mathbf{d}}$ is contained in a $U^{\mathbf{d}}$; i.e., every subgroup of H whose order is a prime number is contained in a cyclic subgroup of H whose order is the square of a prime number. But this shows that every subgroup of H whose order is a prime number is of order q. There exist therefore exactly $p + 1$ proper subgroups of $Z^{\mathbf{d}}$ and all of them are of order q. If we use a theorem on the structure of the lattice of subgroups,[16] it follows that $p = q$, and that $Z^{\mathbf{d}}$ is a direct product of two cyclic groups of order p and is therefore Abelian.

If U is generated by u and V by v, then $U^{\mathbf{d}}$ and $V^{\mathbf{d}}$ are cyclic subgroups of order p^2 in H; and their p-th powers are independent subgroups of order p in $Z^{\mathbf{d}}$. Since $H/Z^{\mathbf{d}}$ is a cyclic group of order p, and since $Z^{\mathbf{d}}$ is Abelian, this implies that H is Abelian and that $H^p = Z^{\mathbf{d}}$. But the statements: H is an Abelian group, $Z^{\mathbf{d}}$ is a direct product of two cyclic groups of order p, $H/Z^{\mathbf{d}}$ is of order p, and $H^p = Z^{\mathbf{d}}$ are incompatible so that the hypothesis of the existence of a dual of G has led to a contradiction.

II. Assume now in general that G is a group which is generated by elements of order p, is not Abelian and has Abelian central quotient-group. Then there exists an element c in G which is the commutator of two elements in G and is different from 1. Since $C \leqq Z$, and since both G/Z and Z are direct products of cyclic groups of order p (see footnote 14), it follows that c is of order p, and that Z is the direct product of a group Z^* and of the group generated by c. Z^* is a normal subgroup of G and the commutator subgroup C' of the group $G' =$

[15] Isomorphism, Lemma 5.1.

[16] Isomorphism, Theorem 8.1.

G/Z^* is a cyclic group of order p. If Z' is the central of G', then $C' \leqq Z'$. Furthermore G' is generated by elements of order p.

Z' is the direct product of some subgroup Z^{**} and of C'. Z^{**} is a normal subgroup of G'. If C'' and Z'' are the commutator subgroup and the central of $G'' = G'/Z^{**}$, then $C'' = Z''$ is a cyclic group of order p; and G'' is generated by elements of order p. It follows now from the special case, treated in I, that there does not exist a dual of G''. Since G'' is a quotient-group of G, this implies that there does not exist a dual of G. Thus it follows that the existence of a dual of G implies that G is Abelian; and it is a consequence of Theorem A that G is finite; and this completes the proof.

It will be convenient to use the following notation. If G is any group, denote by $Z(G) = Z_1(G)$ the central of G and by $Z_i(G)$ the subgroup of G which satisfies: $Z_i(G)/Z_{i-1}(G) = Z[G/Z_{i-1}(G)]$.

COROLLARY 5.2. *If the group G is generated by elements of prime order,*[17] *if $G = Z_h(G)$ for some positive integer h, and if there exists a dual of G, then G is Abelian.*

Proof. If G is not Abelian, there exists a normal subgroup N of G so that G/N is not Abelian, but so that the central quotient-group of G/N is Abelian. For either the central quotient-group of G is Abelian, in which case we may choose $N = 1$; or the choice $N = Z_{h-2}(G)$ meets all the requirements, provided h is the smallest positive integer so that $G = Z_h(G)$.

There exists a dual of G/N, since there exists a dual of G. Hence it follows from Theorem 2.1 that all the elements in G/N are of finite order. But a group without elements of infinite order whose central quotient-group is Abelian is the direct product of (finite or infinite) p-groups.[18] Since G/N is not Abelian, at least one of the direct factors in a direct decomposition is not Abelian. Consequently, there exists a normal subgroup M of G which contains N so that (1) G/M is not Abelian, (2) the central quotient-group of G/M is Abelian, (3) G/M is a p-group, i.e., the orders of the elements in G/M are powers of the (fixed) prime number p.

G/M is generated by elements of order p, since G is generated by elements of prime order. There exists a dual of G/M, since there exists a dual of G. Hence it is a consequence of Theorem 5.1 that G/M is Abelian; and this is a contradiction.

The following statement is just a special case of this corollary, since $G = Z_h(G)$ for some positive integer h, if G is a finite p-group.[19]

COROLLARY 5.2*. *If the finite p-group G is generated by elements of order p, and if there exists a dual of G, then G is Abelian.*

COROLLARY 5.3. *If G is a p-group (finite or infinite), if $G = Z_h(G)$ for some positive integer h, and if there exists a dual of G, then the elements of order p in G generate an Abelian subgroup of G.*

[17] These orders may be different prime numbers.

[18] See, e.g., R. Baer, Comp. Math., vol. 1(1934), p. 261, Lemma.

[19] See, e.g., the work of Burnside, cited above.

Note again that the hypothesis $G = Z_h(G)$ is satisfied (and may therefore be omitted), if G is finite.

Proof. Let P be the subgroup of G which is generated by the elements of order p in G; and let $\mathbf{d}$ be a dualism between G and the group H. As P is the join-group of all the subgroups without proper subgroups, $P^{\mathbf{d}}$ is the cross-cut of all the greatest subgroups of H and $P^{\mathbf{d}}$ is therefore a normal subgroup of H. Thus $\mathbf{d}$ induces a dualism between P and $H/P^{\mathbf{d}}$. Furthermore we have $P = Z_h(P)$, since $G = Z_h(G)$. Since finally P is generated by elements of order p, it follows from Corollary 5.2 that P is Abelian.

The following consequence of Corollary 5.2 is nearly obvious.

COROLLARY 5.4. *If $G = Z_h(G)$ for some positive integer h, and if there exists a dual of G, then G/G^p is Abelian, where G^p is the subgroup of G which is generated by the p-th powers of elements in G.*

6. **Proof of Theorem** C. The necessity of the conditions (a) and (b) of Theorem C is an obvious consequence of Theorem E and of the fact that every subgroup and every quotient-group of an Abelian group is Abelian.

Suppose now that G is a finite p-group, satisfying the conditions (a) and (b). Since all groups of order p are Abelian, we may assume that a group is Abelian if it satisfies the conditions (a) and (b) and if its order is smaller than the order of G.

Since G is a finite p-group, its central contains a cyclic subgroup C of the exact order p. G/C satisfies the conditions (a) and (b), since G satisfies them. Since the order of G/C is smaller than the order of G, it follows from the induction-hypothesis that G/C is Abelian.

Let us assume now that G is *not* Abelian. Then it follows that the subgroup C of order p is just the commutator subgroup of G; and it follows from the construction of C that

$$1 < C \leqq Z(G) < G.$$

As $Z(G)$ is a finite Abelian group, and as C is a subgroup of order p, $Z(G)$ is the direct product of two groups Z' and Z'' so that Z' is cyclic and C a subgroup of Z'. Then Z'' is a normal subgroup of G and G/Z'' satisfies the conditions (a) and (b). But G/Z'' is certainly not Abelian. Hence the order of G/Z'' is not smaller than the order of G. This shows that $Z'' = 1$, i.e., $Z(G) = Z'$ and

$Z(G)$ is therefore a cyclic group of some order p^n.

Denote now by P the subgroup of G which is generated by the elements of order p in G. It is a consequence of Corollary 5.3 that

P is Abelian.

Since G is a group with Abelian central quotient-group whose commutator subgroup is of the exact order p, it follows from the well-known properties of such groups that (see footnote 14)

g^p is in $Z(G)$ for every g in G.

It is a consequence of Lemma 4.1 that there exists a subgroup U of G which is chained to G and which does not contain C, since there exists a dual of G, and since $C \neq 1$. The cross-cut of C and of U is 1, since C is of order p. Hence the cross-cut of U and $Z(G)$ is 1, since $Z(G)$ is a cyclic group which contains C as the uniquely determined smallest subgroup. If u is an element in U, then u^p is in $Z(G)$, as has been remarked before; and consequently $u^p = 1$ as an element in the cross-cut of U and $Z(G)$. Thus it has been shown that U is a subgroup of P; and U is chained to P, since U is chained to G. Since P is an Abelian group which is generated by elements of order p, this implies that P/U is a cyclic group of order p, and that P is the direct product of U and C.

P is chained to G, since P is a group between U and G, and since U is chained to G. G/P is Abelian, since $C \leqq P$. Hence *G/P is a cyclic group of order p^n*, since every g^p is in the cyclic group $Z(G)$ of order p^n, and since the cross-cut of P and $Z(G)$ is the cyclic group C of order p.

Suppose now that v is an element in G which generates G mod P. Then there exists an element w in P so that $vwv^{-1}w^{-1} \neq 1$, since G would be Abelian otherwise. The element $c = vwv^{-1}w^{-1}$ generates C, since C is of order p. If s is any element in P, then $vsv^{-1}s^{-1} = c^i$ for some integer i; and it follows from the general properties of groups with Abelian central quotient-group (see footnote 14) that

$$v(sw^{-i})v^{-1}(sw^{-i})^{-1} = vsv^{-1}s^{-1}(vwv^{-1}w^{-1})^{-i} = 1.$$

As sw^{-i} is an element in the Abelian group P, and as v generates G mod P, this implies that sw^{-i} is in the cross-cut C of P and $Z(G)$. The elements c and w form therefore a basis of the Abelian group P and one verifies now that

G is generated by two elements e and f subject to the relations:

$$1 = e^p = f^{p^{n+1}};$$

$$efe^{-1}f^{-1} = f^{p^n} = c \neq 1, \qquad ec = ce.$$

The integer n is certainly positive. Assume now that $p^n = 2$. Then it follows from the well-known properties of groups with Abelian central quotient-group that (see footnote 14)

$$(ef)^2 = ce^2f^2 = 1,$$

and as the elements of order p in G form an Abelian subgroup of G, this would imply that f is also of order 2. But this is impossible and consequently we have the additional conditions:

$$0 < n, \qquad p^n \neq 2.$$

If i, j, k are integers, then

$$(e^if^j)^k = c^{2^{-1}k(k-1)}e^{ik}f^{jk} = e^{ik}f^{jk+2^{-1}k(k-1)p^n}.$$

If j is relatively prime to p, then we are going to prove that there exists an integer k satisfying $(e^if^j)^k = e^{ik}f$. This will be proved, as soon as we have shown that there exists a solution k of the congruence

$$jk + 2^{-1}k(k-1)p^n \equiv 1 \bmod p^{n+1}.$$

There exists certainly an integer k' so that $jk' \equiv 1 \bmod p^{n+1}$. Thus it suffices to prove the existence of an integer x so that, putting $k = k' + xp^n$, we have

$$jx + 2^{-1}(k' + xp^n)(k' + xp^n - 1) \equiv 0 \bmod p.$$

If $p \neq 2$, then it suffices to determine x in such a way that

$$2jx + k'(k'-1) \equiv 0 \bmod p,$$

since n is positive; and this is possible, since $2j$ is prime to p. If $p = 2$, then p^{n-1} is still even, so that it suffices to solve the congruence

$$jx + 2^{-1}k'(k'-1) \equiv 0 \bmod 2;$$

and this is possible, since j is odd. Thus it has been shown that it is possible to determine k in every case in the required way.

The subgroup A of G which is generated by e and f^p is Abelian. A is exactly the subgroup of G which is generated by $Z(G)$ and P. G/A is a cyclic group of order p. If S is a proper subgroup of G which is not contained in A, then S is a cyclic group of order p^{n+1}, since the p-th powers of all the elements in G are contained in $Z(G)$. It follows now from the previous result that there exists in S one and only one element e^if which generates S. This shows that G is subgroup-isomorphic with a direct product of a cyclic group of order p and a cyclic group of order p^{n+1}. But this contradicts condition (b), since G is not Abelian. Thus the assumption that G is a non-Abelian group, satisfying (a) and (b), has led us to a contradiction; and this completes the proof of Theorem C.

7. The following theorem which will be derived from Theorem C contains it as a special case.

THEOREM 7.1. *If $G = Z_h(G)$ for some positive integer h, and if G satisfies the conditions* (a) *and* (b) *of Theorem* C, *then G is Abelian.*

Proof. Using the same argument as in the proof of Corollary 5.2, we may show that there exists a normal subgroup N of G so that (1) G/N is not Abelian, (2) the central quotient-group of G/N is Abelian, (3) the orders of all the elements of G/N are powers of the fixed prime number p, if only G itself is not Abelian.

Denote by $G(1)$ the subgroup of G/N which is generated by the elements of order p; and denote by $G(i+1)$ the subgroup of G/N which is generated by those elements in G/N whose order mod $G(i)$ is p. All the $G(i)$ are normal subgroups of G/N.

It is a consequence of Corollary 5.2 and of condition (a) that $G(i)$ as well as

every $G(i + 1)/G(i)$ is a finite Abelian group, and this implies that every $G(i)$ is a finite group.

There exists by condition (a) a dualism $\mathbf{d}$ between G/N and some group H. By the argument used in the proof of Corollary 5.3 one shows that $G(1)^{\mathbf{d}}$ is a characteristic subgroup of H, that $G(2)^{\mathbf{d}}$ is the cross-cut of all the greatest subgroups of the characteristic subgroup $G(1)^{\mathbf{d}}$ of H and is therefore itself a characteristic subgroup of H and so on, so that in particular all the $G(i)^{\mathbf{d}}$ are normal subgroups of H. Thus $\mathbf{d}$ induces a dualism between the finite group $G(i)$ and the quotient-group $H/G(i)^{\mathbf{d}}$.

Thus it follows that each of the groups $G(i)$ is a finite p-group which satisfies the conditions (a) and (b) of Theorem C. Consequently each of the $G(i)$ is Abelian. Since $G(i)$ contains in particular those elements in G/N whose order is a divisor of p^i, there exists corresponding to every pair of elements x and y in G/N an integer i so that $G(i)$ contains both x and y. Hence $xy = yx$, since $G(i)$ is Abelian. Consequently G/N is Abelian and this contradicts (1). The hypothesis that G be not Abelian leads therefore to a contradiction, and this completes the proof.

A few remarks may be added. If the group G is generated by elements of order p, then Corollary 5.2 gives a better result, since there the condition (b) is not needed. If, however, all the elements $\neq 1$ in G are of order p, then it is known (Isomorphism) that the condition (b) is always satisfied.

It has been pointed out in §1 that the condition (b) is not necessary for the existence of a dual. There is some reason to believe that a group G which satisfies $G = Z_h(G)$ for some integer h and which possesses a dual is always subgroup-isomorphic with an Abelian group; and it does not seem improbable that a complete survey of the groups which are subgroup-isomorphic with Abelian groups might furnish the means for a proof of this conjecture.

Appendix: Representation of dualisms. Suppose that the groups G and H are both subgroups of the same group R. If S is a subgroup of G, then denote by S^H the set of all those elements h in H which satisfy $sh = hs$ for every s in S so that S^H is the cross-cut of H and of the centralizer of S in R; and similarly denote by T^G for subgroups T of H the set of all those elements g in G so that $gt = tg$ for every t in T. One verifies easily that $S \leqq S' \leqq G$ implies $S'^H \leqq S^H$, that $S \leqq S^{HG}$ and $S^H = S^{HGH}$ for $S \leqq G$.

The two subgroups G and H of the group R are said to form a *pair in R* if $S = S^{HG}$ for every subgroup S of G and $T = T^{GH}$ for every subgroup T of H. If G and H form a pair in R, then mapping $S \leqq G$ upon $S^H \leqq H$ constitutes a dualism between G and H; and one may say that this dualism is represented by the pairing-operation in R.

Suppose now that G is an Abelian group which possesses a dual. Then G does not contain elements of infinite order and all its primary components $G(p)$ are finite. Denote by $p^{m(p)}$ the maximum order of the elements in $G(p)$, by $Z(p)$ a cyclic group of order $p^{m(p)}$, and by Z the direct product of the groups $Z(p)$.

Any homomorphism h of G into Z which maps almost every $G(p)$ upon the identity is said to be a *character* of G and the group H of all the characters $h(g)$ of G is isomorphic with G. There exists one and essentially only one group R with the following properties:

(1) R contains G, Z and H as subgroups;

(2) R is generated by the elements in G, Z and H;

(3) Z is the central of R;

(4) if g is an element in G and h an element in H, then the commutator $ghg^{-1}h^{-1}$ of g and h is equal to the element $h(g)$ in Z upon which the character h of G maps the element g in G.

It is now a well-known fact in the theory of characters of Abelian groups that G and H form a pair in R, since S^H for $S \leqq G$ consists of exactly those characters of G which map every element in S upon 1, and since T^G for $T \leqq H$ consists of exactly those elements in G which are mapped upon 1 by all the characters in T.

UNIVERSITY OF ILLINOIS.

ALGEBRAS DEFINED BY GROUPS WHOSE MEMBERS ARE OF THE FORM $A^x B^y$

By James Byrnie Shaw

This paper is a study of a class of groups of not very complicated structure, considered to be algebras. To make an algebra of a group we must introduce the additive property giving the elements of the group coefficients from a field, in this case the complex field.[1] As an algebra a reducibility appears. It has long been known that an abstract group could be represented in several ways as a linear group, each such representation being practically a matrix, that is, a simple algebra, a quadrate algebra, now called a total matric algebra. It happens thus that a group as an algebra becomes semisimple. The object of the paper is to effect this reduction by simple means. The defining basal units of the algebra, originally called *vids* by C. S. Peirce—a name that should have been retained—, are found by first determining hypernumbers of the algebra which are the partial moduli of the separate integral subalgebras. In the general case of groups this involves solving algebraic equations, but for this class we do not need to do that.

This class of groups exemplifies almost every group property rather simply, and also many properties of algebras. In particular, when there is a central of the group different from ordinary unity, we can extend the field and reduce the order of the algebra. Interesting examples of some of the recent work in algebras can be formed.

I. The groups

The most elementary groups are the cyclic, defined by a single generator A with $A^n = 1$. We may take as next most elementary those defined by two generators A, B with every member of the form $A^x B^y$ and with the relations

$$A^a = 1 = B^b, \qquad A^i = B^j, \qquad BA = A^q B^r, \qquad ci \equiv 0\ (a), \qquad cj \equiv 0\ (b).$$

For such groups we have the

Theorem.

$$B^x A^y = A^{yq^x} B^{xr^y} \qquad (x = 1, 2, \cdots, a;\ y = 1, 2, \cdots, j);$$

$$q^c \equiv 1\ (a), \qquad r^d \equiv 1\ (b), \qquad a = dg, \qquad b = ch, \qquad q = 1 + \omega d, \qquad r = 1 + \mu c.$$

Synthetic proof. Cayley showed (Amer. Jour. Math., vol. 1(1878), pp. 174–176) that abstract groups could be represented by assigning to each generator a

Received May 23, 1939.

[1] See Wedderburn, *Lectures on Matrices*, American Mathematical Society Colloquium Publications, vol. 17, 1934, pp. 167–168.

color and constructing a linkage with the assigned colors, according to the defining equations. Traversing a link from one vertex to another represents a member of the group. Reversing the direction represents the inverse. The links may all be straight. The set consisting of the generator A and its powers is represented by a polygon of a sides, regular in most cases (though this is not essential) and convex or star-shaped. The automorphs of the cyclic group of A will be the original polygon and its possible irreducible stars, and their reverses. A star is given by connecting as adjacent vertices the points reached in the original polygon by A^q, A^{2q}, $\cdots$, where q is prime to a, that is, $(a, q) = 1$. No polygon can be separable. If these models are executed in space of three or more dimensions, very interesting polyhedra appear. For example, the octahedral group is represented by the edges of a snub cube, with proper colors and directions, and the icosahedral group by the edges of a snub dodecahedron.

Models for the groups under consideration are quite simple. To construct one we start with a regular convex polygon of a sides for the generator A. The generator B will furnish polygons of b sides, one attached at each vertex of the polygon for A. It will not connect again with the polygon for A unless we have a relation $A^i = B^j$, in which case the j vertex of B must be the i vertex of A. Corresponding to the members of the group there will then be ab vertices or aj vertices, since we may take j as a divisor of b. It is important to notice that *it must make no difference which vertex we use as the initial point*, the law of construction being the same from any vertex. Now since we have $BA = A^qB^r$, or $B^rA^{-1} = A^{-q}B$, we will proceed backward on the A polygon from the initial point representing identity or 1, attaching polygons as follows: at identity the original convex polygon for B; at A^{-1} the automorph given by B^r. This must be an automorph since it must represent the cyclic group of B. The vertex for B^rA^{-1} is adjacent to that for A^{-1}. But we must reach the same vertex by $A^{-q}B$, that is, we proceed on the B polygon to the first vertex, then on an A polygon to our point by A^{-q}. This second polygon for A must be an automorph. As it makes no difference where we start, we see that when we proceed along the A polygon backwards, each B polygon is the r automorph of the preceding, and along the polygon for B each A polygon is the $-q$ automorph of the preceding. Any vertex may now be reached by two routes giving

$$B^{zr^y}A^{-y} = A^{-yq^z}B^z \quad \text{or} \quad B^zA^y = A^{yq^z}B^{zr^y}.$$

To make plain what we do, consider two instances:

$$A^{10} = 1 = B^8, \qquad BA = A^3B^5.$$

Construct a regular decagon and at the vertices attach octagons as follows: at the first vertex a regular convex octagon, at the next a star octagon, then a convex, then a star, and so on to the beginning. Connect the points of the octagons by decagons as follows: next to the original decagon the automorph for A^3, then the same automorph for this, which is actually the original decagon reversed, then this automorph again for the reverse which will be the star reversed, then the original decagon, and so on a second time. We have now a

linkage of 80 vertices properly connected. The colors for A and B are different. The second instance is

$$A^8 = 1 = B^4, \qquad BA = A^3B, \qquad A^4 = B^2.$$

This group is of order sixteen. Its model consists of a regular octagon for the top, a star octagon for the bottom. In a different color the polygons for B will go directly from a top point to the vertex immediately below, then up to the vertex in the top diametrically opposite the first, thence down, and back to the beginning. The second equation could be

$$BA = A^7B^3$$

giving the same model.

Since the successive polygons are automorphs,

$$(q, a) = 1 = (r, b).$$

If $A^i = B^j$, we change the second equation so that $r < j$. For some least value c we must have $q^c \equiv 1\ (a)$. The number c is a divisor of $\lambda(a)$, a well-known function of an integer from the theory of integers. It is a divisor of the totient of a, $\varphi(a)$. It is the least common multiple of the totients of the prime factors of a to their highest powers, except that for powers of 2 greater than 2^2 we take one-half the totient. The numbers relatively prime to a satisfy congruences $p^x \equiv 1\ (a)$, where x is a divisor of $\lambda(a)$, and when it equals $\lambda(a)$ we call p a primitive root. We notice $b = ch$. Likewise we have $r^d \equiv 1\ (b)$, $a = dg$, and d is a divisor of $\lambda(b)$. These restrictions limit the possible groups of this type for a given order. On account of these relations if we follow along the A polygon, after d steps the automorph of B is the original polygon. That is,

$$B^rA^d = A^{qd}B^r = A^{dq^r}B^r.$$

Hence $d(q^r - 1) \equiv 0\ (a)$ or $q^{r-1} \equiv 1\ (g)$ and

$$r = 1 + \mu c.$$

Similarly,

$$q = 1 + \omega d.$$

We represent A^d by R and B^c by S. $BR = R^qB$, $AS = S^rA$, $RS = SR$. In some cases we will consider, the values of c and d derived from the first three equations are too large when we also have $A^i = B^j$.

Analytic proof. We will have the general relation

$$B^xA^y = A^{f_{xy}}B^{g_{yx}},$$

where the functions f and g may be taken as polynomials since there is always a finite number of values of the arguments and of the function values. The latter are always the former in different permutations. If there is a relation $A^i = B^j$, we may reduce the exponents so that either i is a divisor of a or j of b, or sometimes both cases happen. The exponents in the relation above can also be

reduced so that of A will be less than i or that of B less than j. We shall suppose the latter is the case. Then x will run from 1 to j and y from 1 to a. We now make use of the associative property, first on the left, then on the right, and find

$$B^z B^x A^y = A^{f_{x+z}y} B^{g_y(x+z)} = B^z A^{f_x y} B^{g_y x} = A^{f_z f_x y} B^{g_y x + g_{f_x y} z}.$$

We must then have the congruences

$$f_z f_x y \equiv f_{x+z} y \qquad (a),$$

$$g_y(x + z) - g_y x \equiv g_{f_x y} z \qquad (j).$$

These hold for every x, y, z. The first shows that $f(\quad)$ is an iterative function and $f_x \equiv f^x$. A similar statement holds for $g(\quad)$ and $g_x \equiv g^x$. The second congruence shows that $g_w z$ is a multiple of z, and also $f_w z$ is a multiple of z. We write the corresponding congruence for f_x :

(1) $$f_x(z + y) - f_x y \equiv f_{g_y x} z \qquad (a).$$

We let accents indicate the formal derivative, algebraically defined, and from $f_x f_y z \equiv f_{x+y} z$ we have

$$f'_x(f_y z) \cdot f'_y z \equiv f'_{x+y} z \qquad (a).$$

Let $z \equiv 0$ and notice that $f_x 0 \equiv 0$, getting

$$f'_x 0 \cdot f'_y 0 \equiv f'_{x+y} 0 \qquad \text{for all } x,\ y,$$

whence

$$f'_x 0 \equiv (f'_1 0)^x \equiv q^x.$$

From the relation (1) above

$$f'_x(z + y) \equiv f'_{g_y x} z, \qquad f''_x(z + y) \equiv f''_{g_y x} z.$$

Making $z = 0$, we get

$$f'_x y \equiv f'_{g_y x} 0 \equiv q^{g_y x}.$$

Again we have

$$f''_x(f_y z) \cdot (f'_y z)^2 + f'_x(f_y z) \cdot f''_y z \equiv f''_{x+y} z.$$

Let $z \equiv 0$ giving

$$f''_x 0 \cdot q^{2y} + f''_y 0 \cdot q^x \equiv f''_{x+y} 0.$$

Starting with $x = y = 1$ and proceeding to $2, 3, \cdots, w$, we find

$$f''_w 0 \equiv f''_1 0 \cdot (q^{w-1} + q^w + \cdots + q^{2w-2}),$$

$$f''_x z \equiv f''_1 0 \cdot (q^{g_z x - 1} + q^{g_z x} + \cdots).$$

Change z to $f_y z$ and reduce by the relations above and we have

$$f''_x(f_y z) \equiv f''_1 0 \cdot q^{-g_z y} \cdot (q^{g_z (x+y) - 1} + \cdots)$$

$$\equiv q^{-g_z y} f''_{x+y} z.$$

Set $z = 0$, $x = 0$, and $0 \equiv q^{-y} \cdot f_y''0$, hence $f_y''0 \equiv 0\ (a)$. Hence $f_z''z \equiv 0$, and all succeeding derivatives are zero. We must have then

$$f_x z \equiv z f_x' 0 \equiv z q^x \equiv z q^{g_y x}.$$

We can carry through a similar procedure for $g(\quad)$ and finally we have the theorem

$$B^x A^y = A^{yq^x} B^{xr^y}.$$

We notice that $q^x \equiv q^{xr^y}$ for all x, y, and this can be written simply as $q^r \equiv q\ (a)$. Likewise $r^q \equiv r\ (j)$. We must have $(q, a) = 1$, $(r, b) = 1$, since all powers of A must appear on the right, as well as all powers of B.

For some c equal to j or less than j, $f_c x \equiv x\ (a)$, hence

$$q^c \equiv 1 \quad (a).$$

Since $q^{r-1} \equiv 1\ (a)$ we must have if $q \not\equiv 1$, $r \equiv 1 + \mu c$. Likewise for some d less than a, we may take $a \geqq b$ and we have $g_d x \equiv x\ (j)$. Hence $r^d \equiv 1\ (j)$. Since $r^{q-1} \equiv 1\ (j)$ if $r \not\equiv 1$, $q \equiv 1 + \omega d$. If $q = 1$, $BA = AB^r$, and if $r = 1$, $BA = A^q B$; so that these relations hold in any case. If $A^d = R$ and $B^c = S$, $BR = R^q B$, $SA = AS^r$, $RS = SR$. Consequently in any case

$$BA = ABR^{\omega'} S^{\mu}.$$

R at least is not 1 so that the quotient group of R is $DC = CD^{\mu}$. R, S define an Abelian subgroup whose quotient group is Abelian. R is the lowest power of A for which $BR = R^q B$, and S is the lowest power of B for which $SA = AS^{\mu}$. If c and d are determined without the relation $A^i = B^j$, when this relation is adjoined a new c and d may be found in some cases less than the original c and d. There are other methods of proving these relations and the general theorem, but none is direct as is the one given. They demand the consideration of special cases.

A considerable list of numbers play important parts in the development to follow and we list them here. $\lambda(N)$ is the Cauchy maximum indicator of N.

$$a = dg, \quad R = A^d, \quad R^g = 1, \quad r^d \equiv 1\ (b), \quad \lambda(b) = b'd, \quad i = dg'.$$

$$b = ch, \quad S = B^c, \quad S^h = 1, \quad q^c \equiv 1\ (a), \quad \lambda(a) = a'c, \quad j = ch'.$$

$$q = 1 + \omega d, \quad g = l\eta, \quad \omega = k\eta, \quad (k, l) = 1, \quad d = s\kappa, \quad l = u\kappa,$$

$$(s, u) = 1 = (k, u) = (k, \kappa), \quad \alpha = \eta\kappa.$$

$$r = 1 + \mu c, \quad h = n\theta, \quad \mu = m\theta, \quad (m, n) = 1, \quad c = t\pi, \quad n = v\pi,$$

$$(t, v) = 1 = (m, v) = (m, \pi), \quad \beta = \theta\pi.$$

$$a = ls\eta\kappa = us\eta\kappa^2,\ q - 1 = ks\eta\kappa; \quad b = nt\theta\pi = vt\theta\pi^2,\ r - 1 = mt\theta\pi.$$

$$\text{If } q \equiv 1 + xl,\ q^p \equiv 1 + pxl.$$

$$ql \equiv l\ (a), \quad qg \equiv g\ (a), \quad rn \equiv n\ (b), \quad rh \equiv h\ (b).$$

$$q^{x-1} + q^{x-2} + \cdots + 1 \equiv x\ (ks\eta\kappa).$$

If $x = c$, $q^c \equiv 1$, $q^{c-1} + \cdots + 1 \equiv wu\kappa$, $c \equiv 0$ (κ), $c = c_1\kappa$. Also $d \equiv 0$ (π), $d = d_1\pi$. Hence c and d are divisible by $[\kappa, \pi]$ the least common multiple of κ and π, and a and b are each divisible by $\pi\kappa$.

If $q \equiv 1$ (g), $u = 1$, $R^\alpha = 1$, and conversely.

If $r \equiv 1$ (h), $v = 1$, $S^\beta = 1$, and conversely.

$$(A^xB^y)^z \equiv A^{xQ(y)}B^{yR(x)},$$

where

$$Q(y) = 1 + q^y + q^{2y} + \cdots + q^{(z-1)y},$$

$$R(x) = 1 + r^x + r^{2x} + \cdots + r^{(z-1)x}.$$

If we also have $A^i = B^j$, we may reduce q and r by multiples of i or j. We can make either i a divisor of a or j a divisor of b and sometimes both possibilities exist. We may then reduce q to a number less than i, or r to a number less than j. Also the values of c and d that arise from the original equations may now become smaller, remaining, however, divisors of the original values. Examples will be given. We can also define the group by using as generators powers of A and B which are prime to a, b. The values of q and r will then become powers of the old values and no essential properties change. The equation $BA = A^qB^r$ may be written also

$$BA = A^{q-xi}B^{r+xj} = A^{q+yj}B^{r-yj}.$$

An example for all of these numbers is the group defined by

$$A^{720} = B^{504}, \qquad BA = A^{13}B^{37}.$$

We have

$$d = 6,\quad g = 120,\quad \omega = 2,\quad \eta = 2,\quad l = 60,\quad \kappa = 6,\quad u = 10,\quad \alpha = 12,$$

$$c = 2,\quad h = 42,\quad \mu = 3,\quad \theta = 3,\quad n = 4,\quad \pi = 2,\quad v = 7,\quad \beta = 6.$$

Then $B^4A = A^{481}B^{148}$, $BA^3 = A^{39}B^{253}$. If now we have also the relation $A^{120} = B^{84}$, then $B^4A = AB^{484}$, $BA^3 = A^{399}B$, so that d has dropped to 3, and c to 4. The new values are now

$$d = 3,\quad g = 240,\quad \omega = 4,\quad \eta = 4,\quad l = 60,\quad \kappa = 3,\quad u = 20,\quad \alpha = 12,$$

$$c = 4,\quad h = 126,\quad \mu = 9,\quad \theta = 9,\quad n = 14,\quad \pi = 2,\quad v = 7,\quad \beta = 18.$$

Also $BA = A^{613}B^{121} = A^{133}B^{457}$.

II. Invariant subgroups

Let $a = a'a''$, $b = b'b''$. Then a subgroup is generated by $A^{a'}$, $B^{b'}$. If it is invariant,

$$BA^{a'}B^{-1} = A^{a'q}B^{ra'-1} = A^{qa'}B^{xb'}, \qquad A^{-1}B^{b'}A = A^{q^{b'}-1}B^{rb'} = A^{ya'}B^{rb'}.$$

Hence we must have

$$q^{b'} \equiv 1\ (a'), \qquad r^{a'} \equiv 1\ (b').$$

It will also be Abelian if

$$q^{b'} \equiv 1\ (a''), \qquad r^{a'} \equiv 1\ (b'').$$

The quotient group corresponding to the invariant subgroup is defined by

$$C^{a'} = 1 = D^{b'}, \qquad DC = C^{q'}D^{r'}, \qquad q' \equiv q\ (a'), \qquad r' \equiv r\ (b').$$

The quotient group is always a group of the kind we are studying. The importance of invariant subgroups is due to their quotient groups, for the independent algebras of the quotient groups are always included among those of the group. This is of course true for any group whatever. The independent algebras of the invariant subgroups are not usually among those of the group.

A simple example is $A^{18} = 1 = B^{12}, BA = A^5B^7$. Then $\{A^2, B^6\}$ is invariant and Abelian, with quotient group $C^2 = 1 = D^6, DC = CD$, Abelian. $\{A^3, B^2\}$ is invariant and Abelian, with quotient group $C^3 = 1 = D^2, DC = C^2D$, non-Abelian. $\{A, B^2\}$ is invariant and non-Abelian.

For any of these groups $\{A^\kappa, B^\pi\}$ is invariant. There are several others easily found. One of importance is the "central". It contains all the members of the group which are commutative with every member of the group. If its members are given by A^wB^z, then $AA^wB^z = A^{w+1}B^z = A^wB^zA = A^{w+q^z}B^{zr}$. Therefore

$$q^z \equiv 1\ (a), \qquad z(r-1) \equiv 0\ (b), \qquad z \equiv 0\ (c), \qquad z \equiv 0\ (n), \qquad z \equiv 0\ (vc).$$

Likewise $w \equiv 0\ (ud)$. The central is therefore generated by R^u, S^v. It is Abelian and invariant, but its quotient group may not be Abelian. The generators of the central may be considered algebraically as defining numbers which are commutative with all numbers of the algebra, and the order of the algebra can be reduced by introducing these numbers. As an example $A^{18} = 1 = B^9, BA = A^7B^4$. If we take $A^3 = R, B^2 = S, BA = R^2SAB$, where $R^6 = 1 = S^3$, the order drops to 6. The earliest instance of this procedure is found in quaternions, defined by

$$i^2 = -1 = j^2, \qquad ji = -ij.$$

An important invariant subgroup is that generated by R, S. It is Abelian. If we set $R^\omega = V, S^\mu = W$, then a proper power of V will give R^γ, and a proper power of W will give S^θ. These are the lowest powers of R and S that appear in the commutators of the group. The commutator subgroup is invariant and of much importance. We can take as the general commutator

$$C_{x,y} = A^{-x}B^yA^xB^{-y} = A^{x(q^y-1)}B^{y(r^x-1)},$$

where x, y are to take all values. We may express $C_{x,y}$ as $V^{xY}W^{yX}$, where $Y \equiv q^{y-1} + \cdots + 1$ and $X \equiv r^{x-1} + \cdots + 1$. We note that $V^l = 1 = W^n$. If we take $x \equiv d$, then $C_{d,1} = V^d = V^{s\kappa}$. Since $(s, l) = 1$, a proper power of V reduces to V^κ. Also a proper power of W reduces to W^π. $C_{11} = VW$. The κ and π powers of this enable us to find V^π, W^κ. From these we may find V^ζ, W^ζ, where $(\kappa, \pi) = \zeta$. No lower powers of V, W can be found from the

commutators. Hence the commutator subgroup is generated by V^{ζ}, W^{ζ}, together with VW. If the first two reduce to 1, or if they are powers of VW (when $(u\kappa', v\pi') = 1$), the commutators are powers of VW. If we raise C_{11} to the power yX and divide the general commutator, we have V to the power $xY - yX = xy + \varphi\kappa - xy - \psi\pi$ which can give no lower power than ζ. As an example, the commutator subgroup of $A^{84} = 1 = B^{42}$, $BA = A^{19}B^{19}$, is generated by V, W and contains members which are not commutators. For $A^{16} = 1 = B^{16}$, $BA = A^{5}B^{13}$ the commutator group is generated by $A^{4}B^{12}$.

The commutators have some important relations which we list:

$$C_{xy} = C_{-xq^y,-yr^x} = C^{-1}_{xq^y,-y} = C^{-1}_{-x,yr^x},$$

$$C^{-1}_{xy} = C^{-1}_{-xq^y,-yr^x} = C_{xq^y,-y} = C_{-x,yr^x},$$

$$C_{x,yr^w}C_{wy} = C_{x+w,y} = C_{w,yr^x}C_{xy},$$

$$C_{xq^s,y}C_{xs} = C_{x,y+s} = C_{xq^y,s}C_{xy},$$

$$C_{xy}C_{xq^y,-y} = 1 = C_{xy}C_{-x,yr^x},$$

$$C_{xy}C_{xq^y,-y} = C_{x,0} = 1,$$

$$C_{0,y} = 1, \qquad C_{-x,-y} = C_{du-x,cv-y}.$$

The commutators may be arranged in a table, by placing C_{xy} in the x row and the y column. We shall represent the sum of all the members of a row by row (x), and the sum of all in a column by col (y). That is,

$$\text{row}\ (x) = \sum_y C_{xy}, \qquad \text{col}\ (y) = \sum_x C_{xy}.$$

The table is easily constructed. We first calculate $q - 1, q^2 - 1, \cdots, q^c - 1 \equiv 0\ (a)$, and $r - 1, r^2 - 1, \cdots, r^d - 1 \equiv 0\ (b)$. We now write a top row of powers of A given by the q set, repeated v times. Then we write in the first column at the left powers of B given by the r set repeated u times. The table is then finished by writing into rows downward the double, triple, etc., powers of A in the first row, and in the columns proceeding to the right the double, triple, etc., powers of B in the first column. The cv column and du row will all come out 1. This is all of the table needed as this part would be repeated β times to the right and α times downward. We may now find the sum of all the members of a class of conjugates. The class containing B^y will have as its sum B^y col (y). The class containing A^x will give A^x row (x). The class containing A^xB^y will turn out to be the product $(A^x)(B^y)$. In each set there may be repeated members each the same number of times. We now have the theorems:

$$A^{-w}B^yA^w = C_{wy}B^y = B^yC_{wq^{-y},y} = B^yC_{-w,-yr^w}.$$

$$\sum_w A^{-w}B^yA^w = \text{col}\ (y)\cdot B^y = B^y\cdot\text{col}\ (y) = (B^y).$$

$$B^z A^x B^{-z} = A^x C_{xz} = C_{z,zr-z} A^x = C_{-zq^z,-z} A^x.$$

$$\sum_z B^z A^x B^{-z} = A^x \cdot \text{row } (x) = \text{row } (x) \cdot A^x = (A^x).$$

$$A^{-w} C_{xy} A^w = C_{x,yr^w}, \quad A^{-w} \text{ col } (y) A^w = \text{col } (yr^w), \quad A^{-w} \text{ row } (x) A^w = \text{row } (x).$$

$$B^z C_{xy} B^{-z} = C_{xq^z,y}, \quad B^z \text{ col } (y) B^{-z} = \text{col } (y), \quad B^z \text{ row } (x) B^{-z} = \text{row } (xq^z).$$

$$A^{-w} C_{xy} B^y A^w = C_{x,yr^w} C_{wy} B^y = C_{x+w,y} B^y \cdot A^{-w}(B^y) A^w = \text{col } (y) \cdot B^y.$$

$$\sum A^{-w}(B^y) A^w = a \text{ col } (y) B^y.$$

$$B^z A^x C_{xy} B^{-z} = A^x C_{xq^z,y} C_{xz} = A^x C_{z,y+z} \cdot B^z (A^x) B^{-z} = \text{row } (x) \cdot A^x.$$

$$\sum B^z (A^x) B^{-z} = b \text{ row } (x) A^x.$$

$$\sum_z \sum_w B^z A^{-w} A^x B^y A^w B^{-z} = \text{row } (x) \cdot \text{col } (yr^{-x}) \cdot A^x B^y$$

$$= A^x \text{ row } (x) \cdot \text{col } (y) B^y = (A^x)(B^y).$$

$$\sum_w \sum_z A^{-w} B^z B^y A^x B^{-z} A^w = \text{col } (y) \cdot \text{row } (xq^y) \cdot B^y A^x \equiv (B^y)(A^x).$$

The sum of all members of a class of conjugates is obviously commutative with every member of the group and therefore with every hypernumber constructed from the group. Each sum is the sum of numerical multiples of certain mutually nilfactorial partial moduli each belonging to one subalgebra. These are determined in the theory of group characteristics by making a multiplication table for the various (A^x), (B^y), giving a commutative algebra; then this is reduced to a new form in terms of independent mutually nilfactorial idempotents, in number as many as the number of distinct classes of conjugates. The process is tedious at the best and often impracticable. We will push the analysis a little further, however, and arrive at a simpler method. In fact, we derive the partial moduli of the subalgebras directly and do not have to solve troublesome algebraic equations.

III. **Determination of the algebras**[2]

The problem of reducing the group to an algebra is to find the independent subalgebras into which the group may be reduced, and to express the generators A and B in terms of the units of these subalgebras. It is already known that these subalgebras are quadrates. That is, each consists of p^2 independent units, of the form e_{xy}, where we have

$$e_{xy} e_{wz} = \delta_{yw} e_{xz},$$

where $\delta_{yw} = 1$ if $y = w$, $\delta_{yw} = 0$ if $y \neq w$.

The order of the group is the sum of a set of squares, of which at least one in any case is unity, but for the groups we are considering there will be more than one of order unity. Only perfect groups (which include *non-cyclic* simple

[2] Cf. Trans. Amer. Math. Soc., vol. 5(1904), pp. 326–342. See also *Synopsis of Linear Associative Algebra*, Carnegie Publication no. 78, 1907, XXVI.

groups) have a single subalgebra of order unity. Since the commutator subgroup of these groups is always of order less than that of the group, its quotient group will be of order greater than unity, and there will be as many algebras of order 1^2 as the order of this quotient group.

The hypernumber A has the equation $A^a - 1 = 0$, hence its "latent roots" are the a-th roots of unity and are all distinct. We may therefore express A in the form $(\sigma^a = 1)$

$$A = \sum_{x=1}^{a} \sigma^x \kappa_x, \qquad \kappa_x^2 = \kappa_x,\ \kappa_x \kappa_y = 0 \text{ if } x \neq y.$$

We can also express the idempotent units κ_x in terms of powers of A:

$$a\kappa_x = \sum_{y=1}^{a} \sigma^{-xy} A^y.$$

Likewise we may write

$$B = \sum_{x=1}^{b} \tau^x \pi_x, \qquad \pi_x^2 = \pi_x,\ \pi_x \pi_y = 0 \text{ if } x \neq y.$$

$$b\pi_x = \sum_{y=1}^{b} \tau^{-xy} B^y.$$

The hypernumbers κ, π are not the ultimate basal units we desire, but they serve to give the first decomposition of the algebra. We shall see that they consist of hypernumbers from several of the subalgebras. We will consider first a simple initial case,

$$A^a = 1 = B^b, \qquad BA = A^q B.$$

Then we have

$$aB\kappa_x = \sum_{y=1}^{a} \sigma^{-xy} A^{qy} B = \sum_{z=1}^{a} \sigma^{-xq^{-1}z} A^z B = a\kappa_{xq^{-1}} B.$$

Let

$$e_x = \kappa_x + \kappa_{xq} + \cdots + \kappa_{xq^{c-1}};$$

then

$$Be_x = e_x B.$$

The lowest power of B commutative with A is B^c. Hence we set

$$B^c = \sum_{x=1}^{h} \tau^{xc} \sum_{y=1}^{c} \pi_{x+hy}.$$

Let

$$f_x = \pi_x + \pi_{x+h} + \cdots + \pi_{x+(c-1)h}.$$

Then

$$e_x f_y = f_y e_x.$$

The hypernumbers e, f are commutative with every hypernumber of the group. If x is a divisor of a, it may happen that multiplying by q will give x again in fewer than c terms, the number c_1 being a divisor of $c = c_1c'$. Then for the corresponding f we add each time not h but $c'h$. The number of terms in e must be the same as in f. We have this theorem then:

To find the partial moduli of the individual subalgebras of the group, write down the expressions e multiplied by their corresponding f. These will be idempotent, and mutually nilfactorial. The total number of distinct forms $\kappa_p\pi_w$ appearing in all of the moduli will be ab, the order of the group. To express A and B in the different subalgebras, let the units of the t-th algebra be e^t_{xy}, where t is an index. Then

$$A_t = \sum_x \sigma^{tq^x} e^t_{xx}, \qquad B_t = \sum_x \tau^t e^t_{x,\,x+1}.$$

As an example, consider the group $A^{10} = 1 = B^8$, $BA = A^3B$. Then the subscripts for the κ in the different e are

$$\begin{aligned} \text{in } e_1 &: 1, 3, 9, 7; \\ e_2 &: 2, 6, 8, 4; \\ e_3 &: 5; \\ e_4 &: 10. \end{aligned}$$

The corresponding subscripts of π in the f will be

$$\begin{aligned} \text{in } f_1 &: 1, 3, 5, 7; \\ f_2 &: 2, 4, 6, 8; \\ f_3 &: 1, 2, \cdots, 8, \text{ each } f \text{ distinct}; \\ f_4 &: 1, 2, \cdots, 8, \text{ each } f \text{ distinct}. \end{aligned}$$

There are now four partial moduli for subalgebras of order 4^2 each, and sixteen for subalgebras of order 1^2 each, namely,

$$e_1f_1,\ e_1f_2,\ e_2f_1,\ e_2f_2; \qquad \kappa_5\pi_1, \cdots, \kappa_5\pi_8, \kappa_{10}\pi_1, \cdots, \kappa_{10}\pi_8.$$

In each of the subalgebras of order 4^2, we may take A in any way we please so that $A^{10} = 1$. However, they will all be equal to the simplest form, since a transformation of the units e_{xy} would give this form, namely, in the respective subalgebras, the accents t being omitted as no confusion can result:

$$\begin{aligned} A = \sigma e_{11} + \sigma^3 e_{22} + \sigma^9 e_{33} + \sigma^7 e_{44} + \sigma^2 e_{11} + \sigma^6 e_{22} + \sigma^8 e_{33} + \sigma^4 e_{44} + \sigma e_{11} \\ + \sigma^3 e_{22} + \sigma^9 e_{33} + \sigma^7 e_{44} + \sigma^2 e_{11} + \sigma^6 e_{22} + \sigma^8 e_{33} + \sigma^4 e_{44} \end{aligned}$$

and in the sixteen subalgebras of order 1^2:

$$+ \sigma^5 \kappa_5\pi_1 + \sigma^5 \kappa_5\pi_2 + \cdots + \sigma^5 \kappa_5\pi_8 + \kappa_{10}\pi_1 + \cdots + \kappa_{10}\pi_8.$$

Let

$$B = \sum w_{xy} e_{xy}.$$

Then

$$BA = \sum w_{xy} e_{xy} \sigma^{xqy} = A^q B = \sum w_{xy} e_{xy} \sigma^{xq^x+1}.$$

The basal units are independent, so the coefficients must vanish unless $y \equiv x + 1$. The coefficients remaining are arbitrary except that their product must be ± 1. No generality is lost by taking them all equal, as any other form will reduce to this. Hence we write

$$\Sigma = e_{12} + e_{23} + e_{34} + e_{41}.$$

Then

$$B = \tau\Sigma' + \tau^2\Sigma'' + \tau^3\Sigma''' + \tau^6\Sigma^{iv} + \sum \tau^x \kappa_5 \pi_x + \sum \tau^x \kappa_{10} \pi_x.$$

Using the definition we can find now the κ by selecting the coefficients in A of $\sigma, \sigma^2, \cdots, \sigma^{10} = 1$:

$$\kappa_1 = e'_{11} + e'''_{11}, \qquad \kappa_2 = e''_{11} + e^{iv}_{11}, \qquad \kappa_3 = e'_{22} + e'''_{22}, \qquad \cdots,$$

$$\kappa_5 = \kappa_5(\pi_1 + \cdots + \pi_8), \qquad \kappa_{10} = \kappa_{10}(\pi_1 + \cdots + \pi_8).$$

Also in the first subalgebra

$$\pi_1 = \tfrac{1}{4}(1 + \Sigma + \Sigma^2 + \Sigma^3), \quad \pi_2 = 0, \quad \pi_3 = \tfrac{1}{4}(1 + \tau^6\Sigma + \tau^4\Sigma^2 + \tau^2\Sigma^3), \quad \cdots.$$

The commutator subgroup is generated by A^2; hence its quotient group is

$$C^2 = 1 = D^8, \qquad DC = CD.$$

It is Abelian and its algebra consists of the 16 algebras of order 1^2 each. The central is B^4, whose quotient group is

$$C^{10} = 1 = D^4, \qquad DC = C^3D.$$

Its subalgebras consist of 8 algebras of order 1^2 given by even powers of τ, and the 2 of order 4^2 given by even powers of τ.

Well-known groups of this simple character are:

Dihedral, $A^a = 1 = B^2$, $BA = A^{a-1}B$.

If a is odd, there are 2 subalgebras of order 1^2, all others 2^2.

If a is even, there are 4 of order 1^2, all others of order 2^2.

Dicyclic, $A^{2m} = 1 = B^4$, $A^m = B^2$, $BA = A^{-1}B$.

There are 4 algebras of order 1^2, all others of order 2^2.

Metacyclic, $A^p = 1 = B^{p-1}$, $BA = A^qB$, where p is a prime, and q a primitive root.

There are $p - 1$ algebras of order 1^2, and 1 of order $(p-1)^2$.

N-ionic, $A^{n^2} = 1$, $B^n = A^n$, $BA = A^{n+1}B$, order n^3.

These start with the quaternionic group, and continue with groups which represent nonions, sedenions, etc. They may be defined, by using their central, in the form

$$i^n = \rho = j^n, \qquad ji = \rho ij, \qquad \rho^n = 1.$$

The p-th power of $(n + 1)$ will be congruent to $pn + 1$. The subscripts will then be $x, x + nx, x + 2nx, \cdots$, and according as x is prime to n or not, we would find they reduce to $x, x + n, x + 2n, \cdots$ or $x, x + mn, x + 2mn, \cdots$. The equation $A^n = B^n$ gives us the equations

$$\begin{gathered} \kappa_1 + \kappa_{1+n} + \cdots = \pi_1 + \pi_{1+n} + \cdots, \\ \kappa_2 + \kappa_{2+n} + \cdots = \pi_2 + \pi_{2+n} + \cdots, \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \end{gathered}$$

which will reduce the distinct forms. In place of n^4 distinct units in all the subalgebras there will be only n^3. The commutator subgroup is generated by A^n, so its quotient group is

$$C^n = 1 = D^n, \qquad DC = CD.$$

There will be n^2 subalgebras of order 1^2. There will be $\varphi(n)$ of order n^2. Then for each divisor of n, $n = n_1 n_2$, there will be $n_2^2\varphi(n_1)$ of order n_1^2. As an example, $A^{144} = 1$, $B^{12} = A^{12}$, $BA = A^{13}B$. There are 144 of order 1^2, 4 of order 12^2, $2 \cdot 2^2$ of order 6^2, $2 \cdot 3^2$ of order 4^2, $2 \cdot 4^2$ of order 3^2, $1 \cdot 6^2$ of order 2^2, total 242 classes of conjugates and 242 subalgebras. There are 1728 distinct units in the algebras.

For all these cases where $r = 1$, or equally well where $q = 1$, there are several minor theorems or corollaries which are left to the reader.

The models for these groups are very simple in their construction. We make b polygons of a vertices each, starring them in succession according to q; then they are connected by direct polygons of order b. If $A^i = B^j$, the polygons for B go back to the polygons for A. The N-ionic cases above are of this type.

We consider next the general case, starting with the expression for (A^x). Let $\sigma^a = 1 = \tau^b$.

$$\begin{aligned} (A^x) &= \sum_{y=1}^{vc} A^{xq^y} B^{y(r^x-1)} = \sum_{y=1}^{vc} A^{xq^y} \sum_{p=1}^{b} \tau^{py(r^x-1)} \pi_p \\ &= \sum_{w=1}^{a} \kappa_w \sum_{y=1}^{vc} \sigma^{wxq^y} \sum_{p=1}^{n} \tau^{py(r^x-1)} (\pi_p + \pi_{p+n} + \cdots). \end{aligned}$$

Since $q^c \equiv 1 \ (a)$,

$$(A^x) = v \sum_{w=1}^{a} \kappa_w \sum_{y=1}^{c} \sigma^{wxq^y} \sum_{p=1}^{n} \tau^{py(r^x-1)} (\pi_p + \cdots)(1 + \tau^{cp(r^x-1)} + \tau^{2cp(r^x-1)} + \cdots).$$

The last factor vanishes unless

$$p(r^x - 1) \equiv 0 \ (h).$$

Let $b = b'\pi$,

$$p(r^x - 1) \equiv z_x b' \ (b), \qquad z_x \leqq \pi.$$

$$\tau^{p(r^x-1)} = \sigma^{z_x g d_1},$$

$$(A^x) = vh \sum_{1}^{a} \kappa_w \sum_{y=1}^{c} \sigma^{wxq^y + yz_x g d_1} \sum_{1}^{n} (\pi_p + \pi_{p+n} + \cdots).$$

For our purposes numerical coefficients like vh can be dropped. Since $p(r^x - 1) = z_x th = pr(r^x - 1)$, the subscripts for π will contain $p, pr, pr^2, \cdots$.

For (A^d) we have subscripts for κ in sets

$$\sum \kappa_{wq^y+zg}.$$

Since these are permutable with all κ and π, any form (A^x) or (B^x) can be multiplied by them, and their sets for κ cannot contain other subscripts than these nor have more in a set than these. These sets may themselves break up.

If

$$\begin{aligned} w' &\equiv wq^\psi + \varphi z_x gd_1 \quad (a), \\ w'x &\equiv wxq^\psi + \varphi xz_z gd_1 \quad (a) \\ &\equiv wxq^\psi + \psi z_z gd_1 \quad (a). \end{aligned}$$

Hence

$$(\psi - \varphi x)z_x \equiv 0 \ (\pi).$$

Let

$$(z_x, \pi) = \pi_1, \qquad \pi = \pi_1\pi_2, \qquad z_z = z_z'\pi_1, \qquad \text{and} \qquad \psi \equiv \varphi x \ (\pi_2).$$

Hence if $\varphi = 0$, we will have $w' = wq^{\pi_2}, wq^{2\pi_2}, \cdots$. If $\varphi = 1$,

$$w' = wq^x + z_x' xgd_1, \qquad x < \pi.$$

The sets for κ may then have κ with subscripts

$$\begin{aligned} &w, wq^{\pi_2}, wq^{2\pi_2}, \cdots, \\ &wq^x + xz_z' gd_1, wq^{x+\pi_2} + xz_z' gd_1, \cdots, \\ &wq^{2x} + 2xz_z' gd_1, wq^{2x+\pi_2} + 2xz_z' gd_1 + \cdots. \end{aligned}$$

We note that if $\eta \equiv 0 \ (u)$, then $q \equiv 1 \ (l)$ and

$$q^y \equiv 1 + yd\omega \ (a).$$

Hence

$$\sum \sigma^{wxq^y + yz_z gd_1} = 0$$

unless

$$wx(q - 1) \equiv 0 \ (a)$$

or

$$wx \equiv 0 \ (l) \ (\alpha).$$

The general product of π and κ gives

$$ab\pi_j\kappa_i = \sum_{x=1}^{a} \sum_{y=1}^{a} \sum_{w=1}^{b} \sum_{z=1}^{b} \kappa_y \pi_z \sigma^{x(yq^w - i)} \tau^{w(zr^x - j)}.$$

Considering the terms $w' = w + \xi c$, we find they vanish unless

$$zr^x \equiv j \ (h).$$

Likewise

$$yq^{w} \equiv i \ (g).$$

Let

$$y = iq^{-w} + \varphi_w g, \qquad z = jr^{-x} + \psi_x h;$$

then

$$cd\pi_j\kappa_i = \sum \sigma^{x\varphi_w g}\tau^{y\psi_x h}\kappa_{iq^{-w}+\varphi_w g}\pi_{jr^{-x}+\psi_x h}.$$

If $\pi_j\kappa_i = \kappa_i\pi_j$, then

$$\varphi_w g \equiv (q^{w} - 1)i'u, \qquad \psi_x h \equiv (r^{x} - 1)j'v$$

and

$$\sigma^{i'uks\eta\kappa}\tau^{j'vmt\theta\pi} = 1.$$

Let $\rho = \sigma^{\pi s\eta\kappa\kappa'} = \tau^{vt\theta\pi\pi'}$, where $\xi = (\pi, \kappa)$, $\pi = \pi'\xi$, $\kappa = \kappa'\xi$; then

$$\rho^{i'k\pi'+j'm\kappa'} = 1,$$

$$i'k\pi' + j'm\kappa' \equiv 0 \ (\xi).$$

Also

$$j'mvt\theta\pi\kappa \equiv 0 \ (vt\theta\pi^2),$$

$$j'm\kappa' \equiv 0 \ (\pi'),$$

$$j' \equiv 0 \ (\pi'), \qquad j' = j''\pi'.$$

Also

$$i' = i''\kappa'$$

and

$$i''k + j''m \equiv 0 \ (\xi)$$

with

$$i' = i''u\kappa', \qquad j' = j''v\pi'.$$

The number of these forms $\pi_j\kappa_i$ checks with the order of the quotient group of the commutator subgroup.

As a single example consider

$$A^{10} = 1 = B^{8}, \qquad BA = A^{3}B^{5}.$$

Since $5^x - 1$ is a multiple of $h = 2$, $p = 1$ or 2 and $z_x = 1$ if x is odd, $z_x = 0$ if x is even, $gd_1 = 5$. Hence for (A^x) we have the sets

$$(\kappa_1 + \kappa_8 + \kappa_9 + \kappa_2)(\pi_1 + \pi_3 + \pi_5 + \pi_7),$$

$$(\kappa_3 + \kappa_4 + \kappa_7 + \kappa_6)(\pi_1 + \pi_3 + \pi_5 + \pi_7),$$

$$(\kappa_1 + \kappa_3 + \kappa_9 + \kappa_7)(\pi_2 + \pi_4 + \pi_6 + \pi_8),$$

$$(\kappa_2 + \kappa_6 + \kappa_8 + \kappa_4)(\pi_2 + \pi_4 + \pi_6 + \pi_8),$$

$$(\kappa_5 + \kappa_{10})(\pi_1 + \pi_3 + \pi_5 + \pi_7 + \pi_2 + \pi_4 + \pi_6 + \pi_8).$$

The last breaks up into

$$(\kappa_5 + \kappa_{10})(\pi_1 + \pi_5),\ (\kappa_5 + \kappa_{10})(\pi_3 + \pi_7), \text{ and}$$

$$\kappa_5\pi_2\,,\ \kappa_5\pi_4\,,\ \kappa_5\pi_6\,,\ \kappa_5\pi_8\,,\ \kappa_{10}\pi_2\,,\ \kappa_{10}\pi_4\,,\ \kappa_{10}\pi_6\,,\ \kappa_{10}\pi_8\,.$$

We have then four subalgebras of order 4^2, two of order 2^2, and eight of order 1^2.

To express A in each of these we take first in each algebra of order 4^2, sixteen units e_{xy} such that

$$e_{xy}e_{wz} = \delta_{yw}e_{xz}\,.$$

We have similar forms for the algebras of order 2^2. Then the coefficients of A and its units in each algebra are

A:

e_{11}	e_{22}	e_{33}	e_{44}	e_{11}	e_{22}	e_{11}
σ	σ^8	σ^9	σ^2	σ^5	σ^{10}	-1
σ^3	σ^4	σ^7	σ^6	σ^5	σ^{10}	-1
σ	σ^3	σ^9	σ^7			-1
σ^2	σ^4	σ^6	σ^8			-1
						$+1$
						$+1$
						$+1$
						$+1$

To find B we write the general form in each subalgebra and substitute in the equations for A, B, giving, if we set $\lambda = e_{12} + e_{23} + e_{34} + e_{41}$ $(\lambda^4 = 1)$ in each algebra of order 4^2, indicated by accents,

$$B = \tau\lambda' + \tau^3\lambda'' + \tau^2\lambda''' + \tau^4\lambda^{\mathrm{iv}} + \tau(e_{12}^{\mathrm{v}} - e_{21}^{\mathrm{v}}) + \tau^3(e_{12}^{\mathrm{vi}} - e_{21}^{\mathrm{vi}})$$
$$+ \kappa_5\pi_2 + \tau^2\kappa_5\pi_4 + \tau^4\kappa_5\pi_6 + \tau^6\kappa_5\pi_8 + \kappa_{10}\pi_2 + \tau^2\kappa_{10}\pi_4 + \tau^4\kappa_{10}\pi_6 + \tau^6\kappa_{10}\pi_8\,.$$

To complete the treatment we give each κ and π,

$$\kappa_1 = e_{11}' + e_{11}''', \qquad \kappa_2 = e_{44}' + e_{11}^{\mathrm{iv}}, \qquad \kappa_3 = e_{22}'' + e_{22}''', \qquad \kappa_4 = e_{33}'' + e_{44}^{\mathrm{iv}},$$
$$\kappa_5 = e_{11}^{\mathrm{v}} + e_{11}^{\mathrm{vi}} + \kappa_5(\pi_2 + \pi_4 + \pi_6 + \pi_8),$$
$$\kappa_6 = e_{11}'' + e_{22}^{\mathrm{iv}}, \qquad \kappa_7 = e_{44}'' + e_{44}''', \qquad \kappa_8 = e_{22}' + e_{33}^{\mathrm{iv}}, \qquad \kappa_9 = e_{33}' + e_{33}'',$$
$$\kappa_{10} = e_{22}^{\mathrm{v}} + e_{22}^{\mathrm{vi}} + \kappa_{10}(\pi_2 + \pi_4 + \pi_6 + \pi_8).$$
$$4\pi_1 = (1 + \lambda + \lambda^2 + \lambda^3)' + (1 + \tau^2\lambda + \tau^4\lambda^2 + \tau^6\lambda^3)''$$
$$+ 2(e_{11} + e_{22} + \tau^2 e_{12} + \tau^6 e_{21})^{\mathrm{v}}$$
$$4\pi_2 = (1 + \lambda + \lambda^2 + \lambda^3)''' + (1 + \tau^2\lambda + \tau^4\lambda^2 + \tau^6\lambda^3)^{\mathrm{iv}} + 4(\kappa_5 + \kappa_{10})\pi_2\,,$$
$$4\pi_3 = (1 + \tau^6\lambda + \tau^4\lambda^2 + \tau^2\lambda^3)' + (1 + \lambda + \lambda^2 + \lambda^3)''$$
$$+ 2(e_{11} + e_{22} + \tau^6 e_{12} + \tau^2 e_{21})^{\mathrm{v}}$$
$$4\pi_4 = (1 + \tau^6\lambda + \tau^4\lambda^2 + \tau^2\lambda^3)''' + (1 + \lambda + \lambda^2 + \lambda^3)^{\mathrm{iv}} + 4(\kappa_5 + \kappa_{10})\pi_4\,,$$
$$4\pi_5 = (1 - \lambda + \lambda^2 - \lambda^3)' + (1 + \tau^2\lambda + \tau^4\lambda^2 + \tau^6\lambda^3)''$$
$$+ 2(e_{11} + e_{22} + \tau^6 e_{12} + \tau^2 e_{21})^{\mathrm{v}}$$

$$4\pi_6 = (1 - \lambda + \lambda^2 - \lambda^3)''' + (1 + \tau^6\lambda + \tau^4\lambda^2 + \tau^2\lambda^3)^{\text{iv}} + 4(\kappa_5 + \kappa_{10})\pi_8 ,$$

$$4\pi_7 = (1 + \tau^2\lambda + \tau^4\lambda^2 + \tau^6\lambda^3)' + (1 - \lambda + \lambda^2 - \lambda^3)'' + 2(e_{11} + e_{22} + \tau^2 e_{12} + \tau^6 e_{21})^{\text{v}},$$

$$4\pi_8 = (1 + \tau^2\lambda + \tau^4\lambda^2 + \tau^6\lambda^3)''' + (1 - \lambda + \lambda^2 - \lambda^3)^{\text{iv}} + 4(\kappa_5 + \kappa_{10})\pi_8 .$$

These forms exhibit the manner in which the idempotents κ and π cut across the separate subalgebras. The eighty products $\kappa\pi$ are independent hypernumbers of course and represent the algebra. The forms in terms of the eighty units e_{xy} may be chosen in other ways but we merely have an equivalent form for the algebra.

The variety of algebras and their expressions for groups of this type is very great, and groups with quite similar equations may give algebras of quite different characters, owing to the properties of integers. Almost every theorem in groups may be exemplified by groups of this type except the theorems relating to perfect groups.

In case we have an equation $A^i = B^j$, we expand each side in terms of κ, π and equate the expressions for equal coefficients. These sets will then be equal, commutative with all hypernumbers in the algebra, and will reduce the number of forms otherwise appearing. No special complication arises.

UNIVERSITY OF ILLINOIS.

THE JACOBI CONDITION FOR THE DOUBLE INTEGRAL PROBLEM OF THE CALCULUS OF VARIATIONS

By William T. Reid

1. **Introduction.** This paper is concerned with the Jacobi condition for the problem of minimizing a double integral

$$I = \iint_A f(x, y, z, z_x, z_y)\, dx\, dy \tag{1.1}$$

in a class of admissible surfaces $z = z(x, y)$ with fixed values on the boundary C of A. If C is supposed to be a simply closed regular curve, the usual form of the Jacobi condition may be stated as follows: If A_0 is the interior of a simply closed curve C_0, and A_0 is a proper subset of A, then along a minimizing surface E of class C″ there can exist no solution $u(x, y)$ of the accessory (Jacobi) equation such that $u \equiv 0$ on C_0, $u \not\equiv 0$ on A_0, and $|u_x| + |u_y| \not\equiv 0$ on the part of C_0 lying in A. For a minimizing surface of class C′, Schuler [10][1] has stated the above result in terms of the Haar form of the accessory equation for the case in which the boundary C_0 of A_0 lies interior to A. It is to be remarked that for the more classical case in which the minimizing surface is supposed to be of class C″ the assumption $|u_x| + |u_y| \not\equiv 0$ on C_0 is extraneous if C_0 is interior to A and there exists an elementary solution of the Jacobi equation.[2]

The purpose of the present paper is to remove certain factors which are present in the above formulation of the Jacobi condition simply for convenience of proof. We shall allow A to be an arbitrary connected open set with frontier C. The "simply closed regular curve C_0 and its interior A_0, where A_0 is a proper subset of A" is replaced by "a connected open proper subset $\mathfrak{A}$ of A and its frontier $\mathfrak{C}$". Moreover, we do not suppose that the solution u of the accessory equation is defined on the complement of $\mathfrak{A} + \mathfrak{C}$, but simply that u is of class C′ in $\mathfrak{A}$ and such that u, u_x, and u_y approach continuous limit values on $\mathfrak{C}$. Since relatively little is known concerning the qualitative character of the solutions of the type of differential equations with which we are here concerned, this result is a desirable extension of the usual formulation. In §5 it is shown that the assumption $|u_x| + |u_y| \not\equiv 0$ on $\mathfrak{C}$ is extraneous in case $\mathfrak{C}$ belongs entirely to A and there exists an elementary solution of the accessory equation. Finally,

Received June 19, 1939; presented to the American Mathematical Society, October 31, 1936.

[1] Numerals in square brackets refer to the bibliography at the end of this paper.

[2] See Bolza [2], pp. 676–679. It is readily seen that if the interior A_0 of the simply closed curve C_0 is a proper subset of A the proof given by Bolza on pp. 676–678 establishes the form of the Jacobi condition stated above, although Bolza assumes throughout that C_0 is interior to A.

§6 is devoted to some remarks on the extension of the results of this paper to more general multiple integral problems of the calculus of variations.

2. **Formulation of the problem.** Suppose that A is a bounded and connected open region of the xy-plane, and denote the frontier of A by C. It is also supposed that the function $f(x, y, z, p, q)$ is of class C'' in a region $\mathfrak{R}$ of (x, y, z, p, q)-space whose projection in the xy-plane includes $A + C$. A surface

$$z = z(x, y), \qquad (x, y) \text{ on } A, \tag{2.1}$$

will be termed *admissible* if

(i) the function $z(x, y)$ is of class C' on A; moreover, the functions $z(x, y)$, $z_x(x, y)$, $z_y(x, y)$ are continuous on the frontier C of A in the sense that there exist functions $z_0(x, y)$, $z_1(x, y)$, $z_2(x, y)$ which are continuous on $A + C$ and coincide on A with the functions z, z_x, and z_y, respectively (for simplicity of notation, *if* (x_0, y_0) *is a point of* C *we shall write* $z(x_0, y_0)$, $z_x(x_0, y_0)$, $z_y(x_0, y_0)$ *in place of the more accurate notation* $z_0(x_0, y_0)$, $z_1(x_0, y_0)$, $z_2(x_0, y_0)$);[3]

(ii) the values $(x, y, z(x, y), z_x(x, y), z_y(x, y))$, for (x, y) on $A + C$, are in the region $\mathfrak{R}$.

It follows readily that for an admissible surface (2.1) the integral $I[z]$ exists and is finite. It is to be emphasized that the region of integration for $I[z]$ is the open set A. In particular, we are not assuming that the frontier C of A is of two-dimensional measure zero.

Now suppose that E: $z = Z(x, y)$ is a particular admissible surface which renders $I[z]$ a relative minimum in the class of all admissible surfaces z which assume (in the sense of (i)) on C the same values as Z. As a first necessary condition we have that *for every simply closed regular curve* Γ *which lies, together with its interior* Δ, *in* A *the equation*

$$\int_\Gamma f_p\,dy - f_q\,dx = \iint_\Delta f_z\,dx\,dy \tag{2.2}$$

is satisfied, where the arguments of the partial derivatives of f occurring in (2.2) *are the* (x, y, Z, Z_x, Z_y) *belonging to* E.[4]

Let $\zeta(x, y)$ be an admissible variation in the sense that ζ satisfies the above condition (i), and $\zeta \equiv 0$ on C. For sufficiently small values of a, $z = Z + a\zeta$ is an admissible surface which assumes on C the same boundary values as Z. By the usual method of proof it then follows that E must satisfy the further condition

$$I_2[\zeta] \equiv \iint_A 2\omega(x, y, \zeta, \zeta_x, \zeta_y)\,dx\,dy \geqq 0, \tag{2.3}$$

[3] It is to be remarked that this hypothesis does not in general insure that $z(x, y)$ can be extended to be of class C' in a neighborhood of $A + C$.

[4] See Coral [5], p. 587. The equation (2.2), although not in the form given by Haar, is equivalent to his system of differential equations.

where

$$2\omega(x, y, \zeta, \pi, \kappa) = f_{pp}\pi^2 + 2f_{pq}\pi\kappa + f_{qq}\kappa^2 + 2f_{pz}\pi\zeta + 2f_{qz}\kappa\zeta + f_{zz}\zeta^2,$$

and the arguments of the second order partial derivatives of f are the set (x, y, Z, Z_x, Z_y) belonging to E.

A function $\zeta = u(x, y)$ which is of class C′ in an open subset $\mathfrak{A}$ of A will be said to satisfy the integral form of the Haar accessory (Jacobi) equation if *for every simply closed regular curve* Γ *which lies, together with its interior* Δ, *in* $\mathfrak{A}$ *the equation*

$$\int_\Gamma \omega_\pi(u)\,dy - \omega_\kappa(u)\,dx = \iint_\Delta \omega_\zeta(u)\,dx\,dy \tag{2.4}$$

is satisfied, where the notation $\omega_\pi(u)$, $\omega_\kappa(u)$, $\omega_\zeta(u)$ *indicates that the arguments of these partial derivatives are the* (x, y, u, u_x, u_y) *belonging to* $u(x, y)$.

The surface E will be called non-singular if at each of its elements (x, y, Z, Z_x, Z_y) the determinant $f_{pp}f_{qq} - f_{pq}^2$ is different from zero. Since along a minimizing surface E it is necessarily true that the quadratic form[5]

$$f_{pp}\pi^2 + 2f_{pq}\pi\kappa + f_{qq}\kappa^2$$

is positive semi-definite, non-singularity implies the positive definiteness of this form.

A function $\zeta(x, y)$ is said to be Lipschitzian with constant M on a given set of points if for arbitrary points (x_1, y_1), (x_2, y_2) of this set $| \zeta(x_1, y_1) - \zeta(x_2, y_2) | \leqq M[(x_1 - x_2)^2 + (y_1 - y_2)^2]^{\frac{1}{2}}$. The following two lemmas will be of use in the following section.

LEMMA 2.1. *Suppose that* $\zeta(x, y)$ *is defined on an open set* $\mathfrak{A}$ *and its frontier* $\mathfrak{C}$, $\zeta \equiv 0$ *on* $\mathfrak{C}$, *and* ζ *is Lipschitzian with constant* M *on* $\mathfrak{A} + \mathfrak{C}$. *If* ζ *is defined as identically zero on the complement of* $\mathfrak{A} + \mathfrak{C}$, *then throughout the entire plane* ζ *is Lipschitzian with constant* M; *moreover, the partial derivatives* ζ_x *and* ζ_y *exist and are equal to zero almost everywhere on the complement of* $\mathfrak{A}$.

That the thus defined function is Lipschitzian throughout the entire plane with constant M is immediate. Consequently, the partial derivatives of ζ, wherever they exist, do not exceed M in absolute value; furthermore, the plane set of points where these derivatives do not exist, which necessarily is a subset of $\mathfrak{A} + \mathfrak{C}$, is of two-dimensional measure zero. In order to show that ζ_x and ζ_y are equal to zero almost everywhere on the complement of $\mathfrak{A}$, suppose first that $\zeta \geqq 0$ on $\mathfrak{A}$. If ζ_x or ζ_y exists at a point of $\mathfrak{C}$, it follows that this derivative must be zero, and the last statement of the lemma is established in this case. In general, we may write $\zeta(x, y)$ as the difference of two non-negative Lipschitzian functions, each of which vanishes on the complement of $\mathfrak{A}$.

[5] Mason's proof is valid if the minimizing surface is supposed to be of class C′. See for example, Bolza [2], pp. 673–675.

LEMMA 2.2. *If $I_2[\zeta] \geqq 0$ in the class of admissible variations defined above, then $I_2[\zeta]$ is also non-negative in the class of functions which vanish on C and are Lipschitzian on $A + C$.*

If ζ is Lipschitzian on $A + C$, clearly $I_2[\zeta]$ exists as a Lebesgue integral. Now suppose that ζ is a function which vanishes on C, is Lipschitzian on $A + C$, and for which $I_2[\zeta] < 0$. Let $\rho_C[x, y]$ denote the distance of the point (x, y) from the frontier C of A. It is readily seen that $\rho_C[x, y]$ is Lipschitzian with constant 1; in particular, the partial derivatives of this function exist almost everywhere and do not exceed 1 in absolute value. For each $k > 0$, denote by A_k the subset of A on which $\rho_C[x, y] \leqq k$; clearly $m(A_k)$, the measure of A_k, approaches zero as $k \to 0$. Now set

$$\zeta_k(x, y) = \zeta(x, y)F_k(\rho_C[x, y]) \text{ on } A,$$

where $F_k(\rho)$ is a function of the real variable ρ which is of class C′ on $-\infty < \rho < +\infty$ and such that

$$F_k(\rho) \equiv \begin{cases} 0 & \text{on } \rho \leqq \tfrac{1}{2}k, \\ 1 & \text{on } \rho \geqq k, \end{cases}$$

$$0 \leqq F_k'(\rho) \leqq \frac{3}{k}, \qquad -\infty < \rho < +\infty.$$

An example of such a function is given by $F_k(\rho) \equiv 0$ on $\rho \leqq \frac{1}{2}k$, $F_k(\rho) \equiv 1$ on $\rho \geqq k$, $F_k(\rho) = (2\rho - k)^2(5k - 4\rho)/k^3$ on $\frac{1}{2}k \leqq \rho \leqq k$. Almost everywhere on A we have

$$\zeta_{kx} = \zeta_x F_k(\rho_C) + \zeta F_k' \rho_{Cx}, \qquad \zeta_{ky} = \zeta_y F_k(\rho_C) + \zeta F_k' \rho_{Cy}.$$

In particular, $\zeta_k \equiv \zeta$ on $A - A_k$. If $\zeta(x, y)$ is Lipschitzian with constant M on $A + C$, then on A_k we have $|\zeta(x, y)| \leqq M\rho_C[x, y] \leqq Mk$, and since $|F_k'(\rho)| \leqq 3/k$, it follows that $|\zeta_{kx}| \leqq 4M$, $|\zeta_{ky}| \leqq 4M$ almost everywhere on A_k. Hence there is a constant K, independent of k, such that almost everywhere on A we have $|2\omega(x, y, \zeta_k, \zeta_{kx}, \zeta_{ky})| \leqq K$. Now

$$I_2[\zeta_k] = I_2[\zeta] - \iint\limits_{A_k} 2\omega(x, y, \zeta, \zeta_x, \zeta_y)\, dx\, dy + \iint\limits_{A_k} 2\omega(x, y, \zeta_k, \zeta_{kx}, \zeta_{ky})\, dx\, dy,$$

and each of the integrals in this formula approaches zero as $k \to 0$ since $m(A_k) \to 0$. Hence for k suitably small the corresponding function ζ_k is such that $I_2[\zeta_k] < 0$. The function ζ_k is still only Lipschitzian, but it has the property that $\zeta_k \equiv 0$ on $A_{\frac{1}{2}k}$, and there is a constant d_0 such that if (x_0, y_0) is an arbitrary point of C the square $|x - x_0| \leqq d_0$, $|y - y_0| \leqq d_0$ contains no point of $A - A_{\frac{1}{2}k}$; that is, ζ_k is identically zero on this square.

Integral means will now be used to define a function which is of class C′, is zero on C, and for which $I_2 < 0$. Understanding that $\zeta_k(x, y)$ is defined as identically zero on the complement of $A + C$, let

$$\zeta_{k;d}(x, y) = \frac{1}{4d^2} \int_{x-d}^{x+d} \int_{y-d}^{y+d} \zeta_k(s, t)\, ds\, dt. \tag{2.5}$$

Since $\zeta_k(x, y)$ is continuous, for each $d > 0$ the function $\zeta_{k;d}(x, y)$ is of class C′, and

$$(2.6)\qquad \begin{aligned} \frac{\partial}{\partial x}\zeta_{k;d}(x, y) &= \frac{1}{4d^2}\int_{y-d}^{y+d}[\zeta_k(x+d, t) - \zeta_k(x-d, t)]\,dt, \\ \frac{\partial}{\partial y}\zeta_{k;d}(x, y) &= \frac{1}{4d^2}\int_{x-d}^{x+d}[\zeta_k(s, y+d) - \zeta_k(s, y-d)]\,ds. \end{aligned}$$

As $\zeta_k(x, y)$ is Lipschitzian, these formulas may also be written

$$\frac{\partial}{\partial x}\zeta_{k;d}(x, y) = \frac{1}{4d^2}\int_{x-d}^{x+d}\int_{y-d}^{y+d}\frac{\partial}{\partial s}\zeta_k(s, t)\,ds\,dt,$$

$$\frac{\partial}{\partial y}\zeta_{k;d}(x, y) = \frac{1}{4d^2}\int_{x-d}^{x+d}\int_{y-d}^{y+d}\frac{\partial}{\partial t}\zeta_k(s, t)\,ds\,dt.$$

In particular, the partial derivatives of $\zeta_{k;d}(x, y)$ do not exceed M in absolute value. The integrand of $I_2[\zeta_{k;d}]$ is, therefore, bounded uniformly with respect to d. As $d \to 0$, $\zeta_{k;d}(x, y) \to \zeta_k(x, y)$ uniformly, while almost everywhere the partial derivatives of $\zeta_{k;d}(x, y)$ approach the corresponding partial derivatives $\zeta_k(x, y)$.[6] Consequently, $I_2[\zeta_{k;d}]$ approaches $I_2[\zeta_k]$ as $d \to 0$, and for d sufficiently small $I_2[\zeta_{k;d}] < 0$. But for $d < d_0$ the function $\zeta_{k;d}$ vanishes on C. Hence the assumption that there is a Lipschitzian function ζ which vanishes on C and for which $I_2[\zeta] < 0$ implies that there is a function of class C′ which vanishes on C, and such that $I_2 < 0$.

COROLLARY. *If $U(x, y)$ is a function which vanishes on C, is Lipschitzian on $A + C$, and for which $I_2[U] = 0$, then for arbitrary functions ζ which vanish on C and are Lipschitzian on $A + C$ we have*

$$I_2[\zeta; U] \equiv \iint_A [\zeta_x\omega_r(U) + \zeta_y\omega_s(U) + \zeta\omega_\zeta(U)]\,dx\,dy = 0.$$

This corollary is immediate since for arbitrary values a,

$$I_2[U + a\zeta] = I_2[U] + 2aI_2[\zeta; U] + a^2I_2[\zeta] \geqq 0.$$

3. An extended Jacobi condition.

The principal result of this paper is the following theorem.

THEOREM 3.1 (EXTENDED JACOBI CONDITION). *If E is a non-singular minimizing surface of class C′ and $\mathfrak{A}$ is a connected open proper subset of A with frontier $\mathfrak{C}$, then along E there can exist no solution $u(x, y)$ of the Haar accessory equation* (2.4) *which is such that $u \not\equiv 0$ on $\mathfrak{A}$, u satisfies condition* (i) *of* §2 *on $\mathfrak{A} + \mathfrak{C}$, $u \equiv 0$ on $\mathfrak{C}$, and $|u_x| + |u_y| \not\equiv 0$ on the part of $\mathfrak{C}$ belonging to A.*

Suppose that for an open subset $\mathfrak{A}$ of A there exists a solution $u(x, y)$ of (2.4) such that $u \not\equiv 0$ on $\mathfrak{A}$, u satisfies (i) on $\mathfrak{A} + \mathfrak{C}$, $u \equiv 0$ on $\mathfrak{C}$, and there are points on the frontier $\mathfrak{C}$ of $\mathfrak{A}$ that are in A and at which u_x and u_y are not both

[6] See, for example, de la Vallée Poussin [6], p. 72.

zero. Let $U(x, y) = u(x, y)$ on $\mathfrak{A}$, $U(x, y) = 0$ elsewhere. It then follows by Lemma 2.1 that U is Lipschitzian and that U_x and U_y are equal to zero almost everywhere on the complement of $\mathfrak{A}$. In particular,

$$
\begin{aligned}
I_2[U] &= \iint_{\mathfrak{A}} 2\omega(x, y, U, U_x, U_y)\, dx\, dy \\
&= \iint_{\mathfrak{A}} [u_x \omega_\pi(u) + u_y \omega_\kappa(u) + u\omega_\zeta(u)]\, dx\, dy.
\end{aligned} \tag{3.1}
$$

We now wish to show that $I_2[U] = 0$. In view of the assumption (i) satisfied by the minimizing surface E and also the solution u of the accessory equation, the functions $\omega_\pi(u)$, $\omega_\kappa(u)$, and $\omega_\zeta(u)$ are continuous on the closed and bounded set $\mathfrak{A} + \mathfrak{C}$. Hence, these functions may be extended to be continuous over the entire plane.[7] We shall denote these extended functions simply by ω_π, ω_κ, and ω_ζ. Let

$$
\omega_{\pi;d}(x, y) = \frac{1}{4d^2} \int_{x-d}^{x+d} \int_{y-d}^{y+d} \omega_\pi(s, t)\, ds\, dt,
$$

the functions $\omega_{\kappa;d}(x, y)$ and $\omega_{\zeta;d}(x, y)$ being defined by the corresponding integral means. It then follows that these functions are of class C′ throughout the (x, y)-plane; in particular,

$$
\begin{aligned}
\frac{\partial}{\partial x} \omega_{\pi;d}(x, y) &= \frac{1}{4d^2} \int_{y-d}^{y+d} [\omega_\pi(x + d, t) - \omega_\pi(x - d, t)]\, dt = \frac{1}{4d^2} \int_{C_{xy;d}} \omega_\pi\, dy, \\
\frac{\partial}{\partial y} \omega_{\kappa;d}(x, y) &= \frac{1}{4d^2} \int_{x-d}^{x+d} [\omega_\kappa(s, y + d) - \omega_\kappa(s, y - d)]\, ds = \frac{-1}{4d^2} \int_{C_{xy;d}} \omega_\kappa\, dx,
\end{aligned} \tag{3.2}
$$

where $C_{xy;d}$ denotes the boundary of the square over which the integral mean is taken.

As $\omega_{\pi;d}(x, y) \to \omega_\pi(x, y)$, $\omega_{\kappa;d}(x, y) \to \omega_\kappa(x, y)$, and $\omega_{\zeta;d}(x, y) \to \omega_\zeta(x, y)$ uniformly for (x, y) in a bounded closed set, it follows that

$$
I_{2;d} = \iint_{\mathfrak{A}} [u_x \omega_{\pi;d}(x, y) + u_y \omega_{\kappa;d}(x, y) + u\omega_{\zeta;d}(x, y)]\, dx\, dy
$$

approaches $I_2[U]$ as $d \to 0$. Since $u \equiv 0$ on $\mathfrak{C}$, by evaluating the first two terms of $I_{2;d}$ as iterated integrals we find

$$
I_{2;d} = \iint_{\mathfrak{A}} u \left[\omega_{\zeta;d} - \frac{\partial}{\partial x} \omega_{\pi;d} - \frac{\partial}{\partial y} \omega_{\kappa;d} \right] dx\, dy. \tag{3.3}
$$

Now consider the integrand of (3.3). If (x, y) is a point of $\mathfrak{A}$ whose distance from $\mathfrak{C}$ is greater than $2^{\frac{1}{2}}d$, the functions $\omega_\pi(x, y)$, $\omega_\kappa(x, y)$, $\omega_\zeta(x, y)$ coincide with $\omega_\pi(u)$, $\omega_\kappa(u)$, $\omega_\zeta(u)$ interior to and on $C_{xy;d}$, and in view of the Haar accessory equation the bracketed term in the integrand is zero. If (x, y) is a point of $\mathfrak{A}$

[7] See, for example, Carathéodory [4], p. 620.

whose distance from $\mathfrak{C}$ does not exceed $2^{\frac{1}{2}}d$, then from the Lipschitzian character of $U(x, y)$ there exists a constant M such that $|u(x, y)| \leqq Md$; if K is a constant such that $\omega_r(x, y)$, $\omega_s(x, y)$ and $\omega_t(x, y)$ do not exceed K in absolute value on the $2^{\frac{1}{2}}d$-neighborhood of $\mathfrak{C}$, then at such a point (x, y)

$$|\omega_{t;d}| \leqq K, \qquad \left|\frac{\partial}{\partial x}\omega_{r;d}\right| \leqq \frac{K}{d}, \qquad \left|\frac{\partial}{\partial y}\omega_{s;d}\right| \leqq \frac{K}{d}.$$

Consequently, there is a constant K_1 such that the integrand of (3.3) does not exceed K_1 in absolute value for all values of d. Since at each point of $\mathfrak{A}$ this integrand approaches zero as $d \to 0$, it follows by a well-known theorem on Lebesgue integrals that $I_{2;d} \to 0$ as $d \to 0$. Hence $I_2[U] = \lim_{d \to 0} I_{2;d} = 0$.

Now let (x_0, y_0) be a point of $\mathfrak{C}$ interior to A at which u_x and u_y are not both zero. It follows from the lemma proved in the following section that there exists a connected open subset $\mathfrak{A}_1$ of $\mathfrak{A}$ with frontier $\mathfrak{C}_1$ such that (x_0, y_0) is a point of $\mathfrak{C}_1$, $u(x, y) \equiv 0$ on $\mathfrak{C}_1$, while $u(x, y) \neq 0$ on $\mathfrak{A}_1$; moreover, if we suppose for definiteness that $u_y \neq 0$ at (x_0, y_0), there is a rectangle R: $|x - x_0| \leqq l$, $|y - y_0| \leqq h$ such that the points of $\mathfrak{C}_1$ in this rectangle are the locus of an equation $y = f(x)$, where $f(x_0) = y_0$, $|f(x) - y_0| < h$ on $|x - x_0| \leqq l$, and $u_y \neq 0$ at all points of $\mathfrak{A}_1 + \mathfrak{C}_1$ in R. For $|x - x_0| \leqq l$ the function $f(x)$ is of class C′ and $f'(x) = -u_x(x, f(x))/u_y(x, f(x))$; furthermore, the curve $y = f(x)$ divides the rectangle R into two parts one of which belongs to $\mathfrak{A}_1$, while the other lies in the complement of $\mathfrak{A}_1 + \mathfrak{C}_1$. If $u(x, y)$ is defined and of class C′ in a neighborhood of (x_0, y_0) this result follows immediately from the implicit function theorem; under the weaker assumptions here made additional proof is required. In other words, we may suppose that $\mathfrak{A}$, $\mathfrak{C}$ satisfy the additional conditions stated above for $\mathfrak{A}_1$, $\mathfrak{C}_1$, *and in the rest of the proof of Theorem* 3.1 *we shall do so.* For simplicity in notation, we write $\mathfrak{A}$ and $\mathfrak{C}$ instead of $\mathfrak{A}_1$ and $\mathfrak{C}_1$.

As $u_y(x, y) \neq 0$ for points of $\mathfrak{A} + \mathfrak{C}$ in R, it follows from the non-singularity of E that for points of $\mathfrak{C}$ in R

$$0 < f_{pp}u_x^2 + 2f_{pq}u_xu_y + f_{qq}u_y^2 = u_x\omega_r(u) + u_y\omega_s(u),$$

and since $f'(x) = -u_x(x, f(x))/u_y(x, f(x))$, the function

$$\omega_r(u)f'(x) - \omega_s(u)$$

is different from zero at the points $(x, f(x))$, $|x - x_0| \leqq l$.

Let $v(x, y)$ be a function of class C′ which is different from zero on the interior of R and which is identically zero on the boundary and exterior of this rectangle. It will now be proved that for such a function v

$$\begin{aligned} I_2[v; U] &= \iint_A [v_x\omega_r(U) + v_y\omega_s(U) + v\omega_t(U)]\,dx\,dy \\ &= \iint_{\mathfrak{A}_R} [v_x\omega_r(u) + v_y\omega_s(u) + v\omega_t(u)]\,dx\,dy, \end{aligned}$$

where $\mathfrak{A}_R$ is the part of R which belongs to $\mathfrak{A}$, is different from zero. If k is a value such that $| f(x) + k - f(x_0) | < h$ on $| x - x_0 | \leqq l$ and the curve $y = f(x) + k$, $| x - x_0 | \leqq l$, lies in $\mathfrak{A}_R$, denote by R_k the subregion of $\mathfrak{A}_R$ bounded by this curve and arcs of the rectangle R. Clearly,

$$I_2[v; U] = \lim_{k \to 0} \iint_{R_k} [v_x \omega_\pi(u) + v_y \omega_\kappa(u) + v\omega_\zeta(u)]\, dx\, dy. \tag{3.4}$$

Moreover, if $\omega_{\pi;d}$, $\omega_{\kappa;d}$ and $\omega_{\zeta;d}$ are defined as above, for each k the integral in (3.4) is the limit as $d \to 0$ of

$$\iint_{R_k} [v_x \omega_{\pi;d} + v_y \omega_{\kappa;d} + v\omega_{\zeta;d}]\, dx\, dy. \tag{3.5}$$

Now there exists a d_0 such that if $d < d_0$ the square $C_{xy;d}$ and its interior corresponding to a point (x, y) of R_k lies in $\mathfrak{A}$. Hence by the Haar accessory equation, we have for $d < d_0$ that the function $\partial\omega_{\pi;d}/\partial x + \partial\omega_{\kappa;d}/\partial y - \omega_{\zeta;d}$ is identically 0 on R_k. Therefore, (3.5) is equal to

$$\iint_{R_k} \left[\frac{\partial}{\partial x}(v\omega_{\pi;d}) + \frac{\partial}{\partial y}(v\omega_{\kappa;d})\right] dx\, dy.$$

As the boundary of R_k is rectifiable and the functions v, $\omega_{\pi;d}$, $\omega_{\kappa;d}$ are of class C' in a neighborhood of R_k and its boundary, Green's Lemma is applicable (see Bray [3]). Moreover, since $v = 0$ on the boundary of R_k except along the curve $y = f(x) + k$, this integral is equal to

$$\pm \int_{x_0 - l}^{x_0 + l} v[\omega_{\pi;d} f'(x) - \omega_{\kappa;d}]\, dx,$$

where the arguments of v, $\omega_{\pi;d}$, and $\omega_{\kappa;d}$ are $(x, f(x) + k)$. First, letting $d \to 0$ to obtain the value of the integral in (3.4), and then letting $k \to 0$, we obtain

$$I_2[v; U] = \pm \int_{x_0 - l}^{x_0 + l} v[\omega_\pi(u) f'(x) - \omega_\kappa(u)]\, dx, \tag{3.6}$$

where the arguments of the integrand are $(x, f(x))$. Since the integrand is different from zero and continuous, we have $I_2[v; U] \neq 0$, whereas $I_2[U] = 0$. On the assumption that the extended Jacobi condition does not hold we have thus arrived at a contradiction to the corollary of Lemma 2.2. This completes the proof of Theorem 3.1.

4. **An auxiliary lemma.** In this section we shall prove the following lemma, the results of which were used in the proof of the preceding section.

LEMMA 4.1. *Suppose that $\mathfrak{A}$ is a bounded connected open set with frontier $\mathfrak{C}$, and that $u(x, y)$ satisfies condition* (i) *of* §2 *on $\mathfrak{A} + \mathfrak{C}$, while $u \equiv 0$, $| u_x | + | u_y | \neq 0$ on $\mathfrak{C}$. If (x_0, y_0) is a point of $\mathfrak{C}$ at which u_x and u_y are not both zero, then there exists a connected open subset $\mathfrak{A}_1$ of $\mathfrak{A}$ with frontier $\mathfrak{C}_1$ such that*

(a) *$u(x, y) \neq 0$ on $\mathfrak{A}_1$, $u(x, y) \equiv 0$ on $\mathfrak{C}_1$, and (x_0, y_0) is a point of $\mathfrak{C}_1$;*

(b) *if we suppose for definiteness that $u_y(x_0, y_0) \neq 0$, there is a rectangle*

R: $| x - x_0 | \leqq l$, $| y - y_0 | \leqq h$ such that the points of $\mathfrak{C}_1$ in this rectangle are the locus of an equation $y = f(x)$, where $f(x_0) = y_0$, $| f(x) - y_0 | < h$ on $| x - x_0 | \leqq l$, and $u_y \neq 0$ at all points of $\mathfrak{A}_1 + \mathfrak{C}_1$ in R. For $| x - x_0 | \leqq l$ the function $f(x)$ is of class C' and $f'(x) = -u_x(x, f(x))/u_y(x, f(x))$; furthermore, the curve $y = f(x)$ divides R into two parts one of which belongs to $\mathfrak{A}_1$, while the other lies in the complement of $\mathfrak{A}_1 + \mathfrak{C}_1$.

Let (x_0, y_0) be a point of $\mathfrak{C}$ at which u_x and u_y are not both zero, and for definiteness suppose that $u_y(x_0, y_0) \neq 0$. We may without loss of generality suppose $u_y(x_0, y_0) < 0$, since otherwise we could consider the function $-u(x, y)$. Then there exist constants $k > 0$, $h > 0$ such that at each point of $\mathfrak{A} + \mathfrak{C}$ belonging to the square D: $| x - x_0 | \leqq h$, $| y - y_0 | \leqq h$ we have $u_y < -k$. Let (x_n, y_n) $(n = 1, 2, \cdots)$ be a sequence of points of $\mathfrak{A}$ lying in D, and tending to (x_0, y_0) as limit, and such that for every n either $u(x_n, y_n) \leqq 0$ or $u(x_n, y_n) \geqq 0$. For definiteness, suppose $u(x_n, y_n) \geqq 0$ $(n = 1, 2, \cdots)$ and consider the segments $S_n : x = x_n$, $y_0 - h \leqq y \leqq y_n$. For fixed n, and y sufficiently close to and less than y_n, the point (x_n, y) belongs to $\mathfrak{A}$ and $u(x_n, y) > 0$. Since $u_y < -k < 0$ on D, it then follows that each point of the segment S_n belongs to $\mathfrak{A}$. Now let y^* be any value satisfying $y_0 - h \leqq y^* < y_0$. For n sufficiently large, $n > n^*$, we have $| y_n - y_0 | < \frac{1}{2}(y_0 - y^*)$ and consequently $u(x_n, y^*) \geqq u(x_n, y_n) + k(y_n - y^*) \geqq \frac{1}{2}k(y_0 - y^*)$. In particular, $u(x_0, y^*) = \lim_{n \to \infty} u(x_n, y^*) \geqq \frac{1}{2}k(y_0 - y^*)$, and (x_0, y^*) is a point of $\mathfrak{A}$. Hence the entire segment $x = x_0$, $y_0 - h \leqq y < y_0$ belongs to $\mathfrak{A}$, and if $\mathfrak{A}_1$ denotes the maximal connected subset of $\mathfrak{A}$ containing this segment and on which $u > 0$, the frontier $\mathfrak{C}_1$ of $\mathfrak{A}_1$ contains the point (x_0, y_0). Clearly $\mathfrak{A}_1$ and $\mathfrak{C}_1$ satisfy the conditions of conclusion (a).

Now by exactly the argument used above it follows that if $(\bar{x}, \bar{y})$ is a point of $\mathfrak{A}_1 + \mathfrak{C}_1$ in D, then the entire segment $x = \bar{x}$, $y_0 - h \leqq y < \bar{y}$ belongs to $\mathfrak{A}_1$. In particular, the point $(x_0, y_0 + h)$ does not belong to $\mathfrak{A}_1 + \mathfrak{C}_1$, while $(x_0, y_0 - h)$ is a point of $\mathfrak{A}_1$. Consequently, there is a value l, $0 < l \leqq h$, such that each point of the segment $| x - x_0 | \leqq l$, $y = y_0 + h$ does not belong to $\mathfrak{A}_1 + \mathfrak{C}_1$, while the segment $| x - x_0 | \leqq l$, $y = y_0 - h$ lies in $\mathfrak{A}_1$. It then follows that on $| x - x_0 | \leqq l$ there is a unique value $f(x)$ such that $| f(x) - y_0 | < h$, $(x, f(x))$ is on $\mathfrak{C}_1$, the segment $f(x) < y \leqq y_0 + h$ is in the complement of $\mathfrak{A}_1 + \mathfrak{C}_1$, and the segment $y_0 - h \leqq y < f(x)$ belongs to $\mathfrak{A}_1$. That is, the points of $\mathfrak{C}_1$ in D are the locus of the equation $y = f(x)$.

Let N be such that $| u_x(x, y) | < N$ for all points of $\mathfrak{A}_1 + \mathfrak{C}_1$ in the rectangle R. It will now be shown that if $0 < t \leqq l$ and $Nt/k < h$, then the entire line segment

$$y = y_0 - Nt/k, \qquad | x - x_0 | \leqq t \tag{4.1}$$

is in $\mathfrak{A}_1$. Otherwise, there would exist a value θ, $-1 \leqq \theta \leqq 1$, such that $u(x_0 + \theta t, y_0 - Nt/k) = 0$, while each point of the segment joining $(x_0, y_0 - Nt/k)$ and $(x_0 + \theta t, y_0 - Nt/k)$, except the second end-point, is in $\mathfrak{A}_1$. Then

$$\begin{aligned} 0 &= [u(x_0 + \theta t, y_0 - Nt/k) - u(x_0, y_0 - Nt/k)] + [u(x_0, y_0 - Nt/k) - u(x_0, y_0)] \\ &= u_x(\bar{x}, y_0 - Nt/k)\theta t - u_y(x_0, \bar{y})Nt/k, \end{aligned}$$

where $\bar{x}$ and $\bar{y}$ are suitable intermediate values. It would then follow that

$$| u_x(\bar{x}, y_0 - Nt/k) | \cdot | \theta | t = | u_y(x_0, \bar{y}) | Nt/k,$$

$$N | \theta | t > kNt/k,$$

that is, $| \theta | > 1$, and this is a contradiction. Hence the entire segment (4.1) is in $\mathfrak{A}_1$.

We shall now show that $y = f(x)$ is continuous and has a derivative at (x_0, y_0). Let $(x_0 + \Delta x, y_0 + \Delta y)$ be a point on this curve such that $0 < | \Delta x | \leqq l$, and $N | \Delta x |/k < h$. Then

$$\begin{aligned} 0 &= u(x_0 + \Delta x, y_0 + \Delta y) - u(x_0, y_0) \\ &= u(x_0 + \Delta x, y_0 + \Delta y) - u(x_0 + \Delta x, y_0 - N | \Delta x |/k) \\ &\qquad + u(x_0 + \Delta x, y_0 - N | \Delta x |/k) - u(x_0, y_0 - N | \Delta x |/k) \\ &\qquad + u(x_0, y_0 - N | \Delta x |/k) - u(x_0, y_0) \\ &= u_y(x_0 + \Delta x, y_1)(\Delta y + N | \Delta x |/k) \\ &\qquad\qquad + u_x(x_1, y_0 - N | \Delta x |/k)\Delta x - u_y(x_0, y_2)N | \Delta x |/k), \end{aligned}$$

where y_1, x_1, and y_2 are suitable intermediate values. Then

$$\frac{\Delta y}{\Delta x} = -\frac{u_x(x_1, y_0 - N | \Delta x |/k) + [N | \Delta x |/(k\Delta x)][u_y(x_0 + \Delta x, y_1) - u_y(x_0, y_2)}{u_y(x_0 + \Delta x, y_1)},$$

and therefore

$$\begin{aligned} f'(x_0) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} &= -u_x(x_0, y_0)/u_y(x_0, y_0) \\ &= -u_x(x_0, f(x_0))/u_y(x_0, f(x_0)). \end{aligned}$$

The same type of argument suffices to show that at each value x on $| x - x_0 | \leqq l$ the function $f(x)$ is continuous and has a derivative given by $-u_x(x, f(x))/u_y(x, f(x))$. Hence $f(x)$ is of class C′ on $| x - x_0 | \leqq l$.

5. **Further results on the Jacobi condition.** Now suppose that for a non-singular minimizing surface E there exists an elementary solution on A for the accessory equation (2.4). By such an elementary solution we shall mean a function $w(x, y; \xi, \eta)$ defined for arbitrary points (x, y), (ξ, η) of A that is of the form

$$(5.1) \qquad w(x, y; \xi, \eta) = \phi(x, y; \xi, \eta) \log [(x - \xi)^2 + (y - \eta)^2]^{\frac{1}{2}} + \psi(x, y; \xi, \eta),$$

where for fixed (ξ, η) the functions ϕ, ψ are of class C′ on A, $\phi(\xi, \eta; \xi, \eta) \neq 0$, and w is a solution of the accessory equation (2.4) on the connected open set obtained by deleting the point (ξ, η) from A. Along a non-singular minimizing surface E which satisfies suitable differentiability properties such an elementary solution is known to exist (see Levi [7], and Lichtenstein [8]). We shall prove the following result.

THEOREM 5.1. *If E is a non-singular minimizing surface for which there exists an elementary solution of the accessory equation on A, and $\mathfrak{A}$ is a connected open subset of A whose frontier $\mathfrak{C}$ also belongs to A, then along E there can exist no solution $u(x, y)$ of the Haar accessory equation* (2.4) *such that $u \not\equiv 0$ on $\mathfrak{A}$, u satisfies condition* (i) *of* §2 *on $\mathfrak{A} + \mathfrak{C}$, and $u \equiv 0$ on $\mathfrak{C}$.*

For suppose that along a non-singular minimizing surface E there did exist a solution $u(x, y)$ of the accessory equations satisfying the conditions stated in the theorem. In the preceding sections we have arrived at a contradiction under the additional assumption that $|u_x| + |u_y| \not\equiv 0$ on $\mathfrak{C}$. Hence there remains the case in which u_x and u_y are both identically zero on $\mathfrak{C}$. In this case the function $U(x, y)$ defined as equal to $u(x, y)$ on $\mathfrak{A}$ and zero elsewhere is of class C' and an admissible variation. Since, as in §3, $I_2[U] = 0$, the function $U(x, y)$ is a minimizing surface for $I_2[\zeta]$ and hence the accessory equation (2.4) is satisfied by U on the set A. Now let (ξ, η) be an arbitrary point of $\mathfrak{A}$, and $w(x, y) = w(x, y; \xi, \eta)$ the corresponding value of the elementary solution. Denote by γ the circle $x = \xi + \rho \cos\theta$, $y = \eta + \rho\sin\theta$, where ρ is chosen so small that γ, together with its interior, is in $\mathfrak{A}$; moreover, let $\mathfrak{A}_\gamma$ denote the subset of $\mathfrak{A}$ exterior to γ, and set

$$(5.2)\qquad \begin{aligned} I_2^{(\gamma)} &= \iint_{\mathfrak{A}_\gamma} [u_x\omega_\pi(w) + u_y\omega_\kappa(w) + u\omega_\zeta(w)]\,dx\,dy \\ &= \iint_{\mathfrak{A}_\gamma} [w_x\omega_\pi(u) + w_y\omega_\kappa(u) + w\omega_\zeta(u)]\,dx\,dy. \end{aligned}$$

As in the preceding sections, let $\omega_{\pi;d}(w)$, $\omega_{\kappa;d}(w)$, and $\omega_{\zeta;d}(w)$ denote the integral means of $\omega_\pi(w)$, $\omega_\kappa(w)$, and $\omega_\zeta(w)$ over the corresponding square with boundary $C_{xy;d}$. Then

$$(5.3)\qquad \begin{aligned} I_2^{(\gamma)} &= \lim_{d\to 0} \iint_{\mathfrak{A}_\gamma} [u_x\omega_{\pi;d}(w) + u_y\omega_{\kappa;d}(w) + u\omega_{\zeta;d}(w)]\,dx\,dy \\ &= \lim_{d\to 0} \left(\iint_{\mathfrak{A}_\gamma} \left[\frac{\partial}{\partial x}\{u\omega_{\pi;d}(w)\} + \frac{\partial}{\partial y}\{u\omega_{\kappa;d}(w)\}\right] dx\,dy \right. \\ &\qquad \left. + \iint_{\mathfrak{A}_\gamma} u\left[\omega_{\zeta;d}(w) - \frac{\partial}{\partial x}\omega_{\pi;d}(w) - \frac{\partial}{\partial y}\omega_{\kappa;d}(w)\right] dx\,dy \right) \\ &= \lim_{d\to 0} - \int_\gamma u[\omega_{\pi;d}(w)\,dy - \omega_{\kappa;d}(w)\,dx] \\ &= -\int_\gamma u[\omega_\pi(w)\,dy - \omega_\kappa(w)\,dx], \end{aligned}$$

where the third equality above results from the fact that $w(x, y)$ satisfies the accessory equation in a neighborhood of $\mathfrak{A}_\gamma + \gamma + \mathfrak{C}$, and also the fact that $u \equiv 0$ on $\mathfrak{C}$.

Now let $\mathfrak{A}_0$ be a connected open subset of A containing $\mathfrak{A} + \mathfrak{C}$ and such that

$\mathfrak{C}_0$, the frontier of $\mathfrak{A}_0$, is in A. Since $U \equiv 0$ except on $\mathfrak{A}$, for d sufficiently small $\omega_{\pi;d}(U)$, $\omega_{\kappa;d}(U)$, and $\omega_{\zeta;d}(U)$ are all zero on $\mathfrak{C}_0$. If $\mathfrak{A}_{0\gamma}$ denotes the subset of $\mathfrak{A}_0$ exterior to γ, then clearly

$$I_2^{(\gamma)} = \iint_{\mathfrak{A}_{0\gamma}} [w_x\omega_\pi(U) + w_y\omega_\kappa(U) + w\omega_\zeta(U)]\,dx\,dy.$$

As $U(x, y)$ is a solution of the accessory equation on A, it then follows by a similar argument that

$$\begin{aligned} I_2^{(\gamma)} &= \lim_{d\to 0} \iint_{\mathfrak{A}_{0\gamma}} [w_x\omega_{\pi;d}(U) + w_y\omega_{\kappa;d}(U) + w\omega_{\zeta;d}(U)]\,dx\,dy \\ &= \lim_{d\to 0} -\int_\gamma w[\omega_{\pi;d}(u)\,dy - \omega_{\kappa;d}(u)\,dx] \qquad (5.4) \\ &= -\int_\gamma w[\omega_\pi(u)\,dy - \omega_\kappa(u)\,dx]. \end{aligned}$$

Consequently,

$$\int_\gamma u[\omega_\pi(w)\,dy - \omega_\kappa(w)\,dx] = \int_\gamma w[\omega_\pi(u)\,dy - \omega_\kappa(u)\,dx]. \qquad (5.5)$$

Since w becomes infinite as $\log \rho$ as $\rho \to 0$, and $dy = \rho\cos\theta\,d\theta$, $dx = -\rho\sin\theta\,d\theta$ along γ, it follows that the right-hand integral in (5.5) approaches zero as $\rho \to 0$. On the other hand, in view of the particular form (5.1) of w, it is an easy calculation to show that as $\rho \to 0$ the left-hand member of (5.5) approaches

$$\pi[f_{pp}(\xi, \eta) + f_{qq}(\xi, \eta)]\phi(\xi, \eta; \xi, \eta)u(\xi, \eta),$$

and the non-singularity condition implies $u(\xi, \eta) = 0$. Since (ξ, η) was chosen as an arbitrary point of $\mathfrak{A}$, it follows that $u(x, y) \equiv 0$ in $\mathfrak{A}$, contrary to hypothesis. Hence Theorem 5.1 is established.

It is to be emphasized that the above proof establishes a necessary condition for a minimizing surface E of class C′, but does not in itself give a proof of a general property of solutions of the accessory equation (2.4). The minimizing property of E was used to show that the above defined function $U(x, y)$ was a solution of (2.4) on the open set A. Analogous to what is proved under more classical conditions (see Bolza [2], p. 679) one would expect to prove the following result for a general equation of the form (2.4).

CONJECTURED RESULT. *Suppose*: (1) *the coefficients of a quadratic form* $2\omega(x, y, \zeta, \zeta_x, \zeta_y)$ *are continuous in* (x, y) *on a bounded open set* $\mathfrak{A}$, *and approach continuous limits on the frontier* $\mathfrak{C}$ *of* $\mathfrak{A}$; (2) *the corresponding equation* (2.4) *is elliptic in the sense that* $\omega_{\pi\pi}\omega_{\kappa\kappa} - \omega_{\pi\kappa}^2 > 0$ *on* $\mathfrak{A} + \mathfrak{C}$; (3) *for each point* (ξ, η) *of* $\mathfrak{A}$ *there exists a corresponding elementary solution of* (2.4) *of the form* (5.1), *such that the functions* ϕ *and* ψ *satisfy condition* (i) *of* §2 *on* $\mathfrak{A} + \mathfrak{C}$. *Then* $u(x, y) \equiv 0$ *is the only solution of* (2.4) *on* $\mathfrak{A}$ *which satisfies condition* (i) *on* $\mathfrak{A} + \mathfrak{C}$, *and for which* u, u_x *and* u_y *are identically zero on* $\mathfrak{C}$.

Clearly this result is true if for u and w relation (5.5) can be established; in turn, this equation holds if (5.3) and (5.4) are provable. Now the above proof of (5.3) did not make use of the minimizing property of E, but it did use the fact that for fixed (ξ, η) in $\mathfrak{A}$ the function $w(x, y) = w(x, y; \xi, \eta)$ was a solution of (2.4) in a neighborhood of $\mathfrak{A} + \mathfrak{C}$. This latter condition can be replaced, however, by the assumption that w satisfies condition (i) on each region $\mathfrak{A}_\gamma$. The functions $\omega_\pi(w)$, $\omega_\kappa(w)$ and $\omega_\zeta(w)$ may then be extended to be continuous throughout the entire plane, and as u is Lipschitzian on $\mathfrak{A} + \mathfrak{C}$ and $u \equiv 0$ on $\mathfrak{C}$, by an argument similar to that used in §3 to show that $I_2[u] = 0$ it follows that

$$\iint_{\mathfrak{A}_\gamma} u \left[\omega_{\zeta;d}(w) - \frac{\partial}{\partial x} \omega_{\pi;d}(w) - \frac{\partial}{\partial y} \omega_{\kappa;d}(w) \right] dx\, dy$$

approaches zero as $d \to 0$. Hence relation (5.3) is true under the conditions listed in the above Conjectured Result.

The author is unable, however, to give a proof of (5.4) under these assumptions. It is possible, however, to establish this latter relation if the further assumption is made:

(4) *If $\mathfrak{A}_k$ is the set of points (x, y) of $\mathfrak{A}$ such that $\rho_{\mathfrak{C}}[x, y] \leqq k$, where $\rho_{\mathfrak{C}}[x, y]$ denotes the distance of (x, y) from $\mathfrak{C}$ and $m(\mathfrak{A}_k)$ is the measure of $\mathfrak{A}_k$, then there is a constant N such that $m(\mathfrak{A}_k)/k \leqq N$ for all values k satisfying* $0 < k \leqq 1$.[8]

For, let $U(x, y) = u(x, y)$ on $\mathfrak{A}$, $U(x, y) = 0$ elsewhere. Then $U(x, y)$ is of class C' in the entire plane, and the functions ω_π, ω_κ and ω_ζ defined as the corresponding $\omega_\pi(u)$, $\omega_\kappa(u)$ and $\omega_\zeta(u)$ on $\mathfrak{A}$, and as zero elsewhere, are continuous in the entire plane. By using the integral means of these functions, and an argument similar to that of §3 in the proof of the relation $I_2[u] = 0$, it follows that if v is Lipschitzian on $\mathfrak{A}_\gamma + \gamma + \mathfrak{C}$, and $v \equiv 0$ on $\mathfrak{C}$, then

$$\iint_{\mathfrak{A}_\gamma} [v_x \omega_\pi(u) + v_y \omega_\kappa(u) + v\omega_\zeta(u)]\, dx\, dy = - \int_\gamma v[\omega_\pi(u)\, dy - \omega_\kappa(u)\, dx]. \tag{5.6}$$

Now for (ξ, η) a point of $\mathfrak{A}$ and $w(x, y) = w(x, y; \xi, \eta)$, set $v_k(x, y) = w(x, y)F_k(\rho_{\mathfrak{C}}[x, y])$, where $F_k(\rho)$ is a function of the sort used in the proof of Lemma 2.2. If k is restricted to values such that the circle γ has no point in common with $\mathfrak{A}_k$, then $v_k = w$ on γ, and as v_k is Lipschitzian on $\mathfrak{A}_\gamma + \gamma + \mathfrak{C}$ and $v_k = 0$ on $\mathfrak{C}$, we have by (5.6) that

$$\begin{aligned} &\iint_{\mathfrak{A}_\gamma} [v_{kx} \omega_\pi(u) + v_{ky} \omega_\kappa(u) + v_k \omega_\zeta(u)]\, dx\, dy \\ &\qquad = - \int_\gamma w[\omega_\pi(u)\, dy - \omega_\kappa(u)\, dx]. \end{aligned} \tag{5.7}$$

Since $v_k = w$ on $\mathfrak{A}_\gamma - \mathfrak{A}_k$, we have in view of (5.7) that

[8] It is readily seen that condition (4) is true if $\mathfrak{C}$ consists of a finite number of isolated points and rectifiable arcs.

$$I_2^{(\gamma)} = \iint_{\mathfrak{A}_k} [w_x\omega_\pi(u) + w_y\omega_\kappa(u) + w\omega_\zeta(u)]\,dx\,dy$$

$$(5.8) \qquad - \iint_{\mathfrak{A}_k} [v_{kx}\omega_\pi(u) + v_{ky}\omega_\kappa(u) + v_k\omega_\zeta(u)]\,dx\,dy$$

$$- \int_\gamma w[\omega_\pi(u)\,dy - \omega_\kappa(u)\,dx].$$

The first integral in (5.8) approaches zero as $k \to 0$ since $m(\mathfrak{A}_k) \to 0$. Now there is an $\epsilon(k)$ which approaches zero as $k \to 0$ and such that $\omega_\pi(u)$, $\omega_\kappa(u)$ and $\omega_\zeta(u)$ do not exceed $\epsilon(k)$ in absolute value on $\mathfrak{A}_k$; moreover, in view of the properties of F_k there is a constant M such that $|v_k| \leqq M$, $|v_{kx}| \leqq M/k$, $|v_{ky}| \leqq M/k$ on $\mathfrak{A}_\gamma$ for $k \leqq 1$. Consequently, if condition (4) is satisfied, the second integral of (5.8) also approaches zero as $k \to 0$; that is, relation (5.4) holds for $I_2^{(\gamma)}$. It has been proved, therefore, that *the conclusion of the above Conjectured Result is true if in addition to hypotheses* (1), (2) *and* (3) *condition* (4) *is also satisfied.*

6. **Remarks.** In the preceding sections we have been concerned with a double integral problem of the calculus of variations involving a single dependent function. The above theorems extend with all the corresponding generality, and with no essential change in argument, to an m-tuple integral problem involving m independent variables $x^1, \cdots, x^m$ and a single dependent function $z = z(x^1, \cdots, x^m)$.

However, for a multiple integral problem, or even a double integral problem, involving more than a single dependent function there is an essential difficulty in carrying through the above type of argument to obtain an extended form of the Jacobi condition with all the corresponding generality. This difficulty lies in the fact that there is no longer a result corresponding to that of Lemma 4.1. For let $u_1(x, y)$, $u_2(x, y)$ be two functions satisfying condition (i) on a bounded open set $\mathfrak{A}$ and its frontier $\mathfrak{C}$, $u_1 \equiv 0 \equiv u_2$ on $\mathfrak{C}$, while u_{1x}, u_{1y}, u_{2x} and u_{2y} are not all identically zero on $\mathfrak{C}$. Suppose, for definiteness, that $u_{1y} \neq 0$ at a point (x_0, y_0) of $\mathfrak{C}$. Then there is a connected subset $\mathfrak{A}_1$ of $\mathfrak{A}$ with frontier $\mathfrak{C}_1$ satisfying conditions (a) and (b) of Lemma 4.1 for the function u_1; there is, however, no assurance that $u_2 \equiv 0$ on $\mathfrak{C}_1$. If, however, we assume that $u_{1y} \neq 0$ at a point (x_0, y_0) of the *outer boundary*[9] of $\mathfrak{A}$, and that near this point $u_1(x, y)$ may be extended to be of class C′, then the usual implicit function theorem is applicable to show that near this point $\mathfrak{C}$ *itself* is representable by an equation of the form $y = f(x)$, and argument similar to that of §2 suffices to prove an extended Jacobi condition for such a problem. The extensibility of u_1 is certainly assured if we initially assume that the solution u_1, u_2 of the accessory equations exists and is of class C′ on a neighborhood of each point of the outer boundary of $\mathfrak{A}$ belonging to A; in particular, if we initially assume that u_1, u_2 exist and are of class C′ on A. Corresponding results for general multiple integral problems involving more

[9] The outer boundary of $\mathfrak{A}$ consists of those points of $\mathfrak{C}$ which are limit points of the complement of $\mathfrak{A} + \mathfrak{C}$.

than a single dependent function are readily proved. For a discussion of the Jacobi condition for multiple integral problems under more classical hypotheses the reader is referred to Raab [9].

BIBLIOGRAPHY

1. G. A. BLISS, *The Calculus of Variations, Multiple Integrals*, Mimeographed notes of lectures delivered summer, 1933, at the University of Chicago.
2. O. BOLZA, *Vorlesungen über Variationsrechnung*, Leipzig, 1909.
3. H. E. BRAY, *Green's Lemma*, Annals of Mathematics, vol. 26(1924-25), pp. 278-286.
4. C. CARATHÉODORY, *Vorlesungen über reelle Funktionen*, Leipzig, 1927.
5. MAX CORAL, *On the necessary conditions for the minimum of a double integral*, this Journal, vol. 3(1937), pp. 585-592.
6. C. J. DE LA VALLÉE POUSSIN, *Intégrales de Lebesgue*, Paris, 1916.
7. E. E. LEVI, *Sulle equazioni lineari totalmente ellittiche alle derivate parziali*, Rendiconti del Circolo Matematico di Palermo, vol. 24(1907), pp. 275-317.
8. L. LICHTENSTEIN, *Neuere Entwicklung der Theorie partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus*, Encyclopädie der Mathematischen Wissenschaften, IIC, 12, pp. 1276-1334.
9. A. W. RAAB, *Jacobi's condition for multiple integral problems of the calculus of variations*, Contributions to the Calculus of Variations, 1931-32, The University of Chicago Press, pp. 445-473.
10. L. SCHULER, *Über die notwendigen Bedingungen beim zweidimensionalen Variationsproblem einfachster Art*, Acta Litterarum ac Scientiarum, vol. 5(1930-32), pp. 154-167.

THE UNIVERSITY OF CHICAGO.

GROUPS AND ABELIAN GROUPS IN TERMS OF NEGATIVE ADDITION AND NEGATION

By B. A. Bernstein

1. **Object.** Let $+$ be the "direct" operation in groups and a' the inverse, or "negative", of a. Let Δ be the operation of "negative addition", given by $a \Delta b = (a + b)'$. The object of this paper is to define groups and Abelian groups in terms of Δ, $'$.

The definitions are postulational. I give a set of independent conditions on a class K, a binary operation Δ, and a unary operation $'$, so that the system $(K, \Delta, ')$ is a group. Similarly for an Abelian group. The postulates are such that the set for groups is embraced in the set for Abelian groups. The use of the operation Δ and the use of a unary operation are, as far as I know, novel in the definition of groups and Abelian groups.[1]

The postulates for groups are P_0-P_5 below; those for Abelian groups are P_0-P_6. Of these, P_0 merely rules out trivial systems—systems in which K is empty or contains but a single element. For unrestricted groups, omit P_0. For merely non-vacuous groups, replace P_0 by: P_0'. *K is not empty.* The postulates and the theorems derived from the postulates will bring out properties of Δ not easily seen when $a \Delta b$ is written in the form $(a + b)'$.

2. **Postulates for groups.** The postulates for groups follow. In P_3-P_5, supply the clause: *whenever the elements involved and their combinations are in K.*

P_0. *K contains at least two distinct elements.*

P_1. *$a \Delta b$ is in K whenever a, b are in K.*

P_2. *a' is in K whenever a is in K.*

P_3. $(a \Delta b)' \Delta c = a \Delta (b \Delta c)'$.

P_4. $(a \Delta b) \Delta a = b$.

P_5. $(a \Delta b)' = b' \Delta a'$.

3. **Theorems. Sufficiency of P_0-P_5 for groups.** Theorems T_1-T_{13} following, derivable from P_0-P_5, will establish the sufficiency of P_0-P_5 for groups.

T_1. $b \Delta (a \Delta b) = a$.[2]

Received June 27, 1939; presented to the American Mathematical Society, April 15, 1939, under the title *Groups and Abelian groups in terms of negative addition.*

[1] For references to postulates for groups and Abelian groups in terms of operations other than the "direct" operation, see footnotes to my *Postulates for abelian groups and fields in terms of non-associative operations*, Trans. Amer. Math. Soc., vol. 43(1938), pp. 1–6. For the first explicit use of an undefined unary operation in a set of postulates, see my paper *Whitehead and Russell's theory of deduction as a mathematical science*, Bull. Amer. Math. Soc., vol. 37(1931), pp. 480–488. The term *unary* was introduced in this paper.

[2] T_1 can evidently be used in place of P_4 in the postulates for groups and Abelian groups.

For, $a = [(a \Delta b) \Delta a] \Delta (a \Delta b) = b \Delta (a \Delta b)$, by P_4, P_4.

T_2. $a'' = a$.

For, $a = (b \Delta c)' \Delta [a \Delta (b \Delta c)'] = (b \Delta c)' \Delta [(a \Delta b)' \Delta c] = b \Delta \{c \Delta [(a \Delta b)' \Delta c]\}' = b \Delta (a \Delta b)''$, by T_1, P_3, P_3, T_1. Therefore,

(i) $$(a \Delta b)'' = a \Delta b,$$

by P_4. Hence, $a'' = [b \Delta (a \Delta b)]'' = b \Delta (a \Delta b) = a$, by T_1, (i), T_1.

T_3. $a \Delta a' = (a \Delta a')'$.

For, $a \Delta a' = a'' \Delta a' = (a \Delta a')'$, by T_2, P_5.

T_4. $a \Delta a' = a' \Delta a$.

For, $a \Delta a' = a \Delta [(a \Delta a) \Delta a]' = [a \Delta (a \Delta a)]' \Delta a = a' \Delta a$, by P_4, P_3, T_1.

T_5. $a \Delta a' = b \Delta b' = b' \Delta b$.

For, $a \Delta a' = a \Delta [(b \Delta a) \Delta b]' = [a \Delta (b \Delta a)]' \Delta b = b' \Delta b = b \Delta b'$, by P_4, P_3, T_1, T_4.

Definition D_1. $z = a \Delta a' = a' \Delta a = (a \Delta a')'$.

T_6. $a \Delta z = z \Delta a = a'$.

Proof: D_1, T_1, P_4.

T_7. $z = z' = z \Delta z$.

For, $z = z \Delta z' = (z \Delta z')' = z' = z \Delta z$, by D_1, T_3, D_1, T_6.

Definition D_2. $a + b = (a \Delta b)'$.

T_8. *$a + b$ is in K.*

Proof: D_2, P_1, P_2.

T_9. $(a + b) + c = a + (b + c)$.

For, $(a + b) + c = [(a \Delta b)' \Delta c]' = [a \Delta (b \Delta c)']' = a + (b + c)$, by D_2, P_3, D_2.

T_{10}. $z + a = a$.

For, $z + a = (z \Delta a)' = a'' = a$, by D_2, T_6, T_2.

T_{11}. $a + z = a$.

For, $a + z = (a \Delta z)' = a'' = a$, by D_2, T_6, T_2.

T_{12}. $a' + a = z$.

For, $a' + a = (a' \Delta a)' = z' = z$, by D_2, D_1, T_7.

T_{13}. $a + a' = z$.

For, $a + a' = (a \Delta a')' = z$, by D_2, D_1.

P_0, T_8-T_{13} make $(K, \Delta, ')$ a (non-trivial) group.

4. **Additional theorems for groups.** Theorems T_{14}-T_{19} following are additional theorems for groups. The proofs are left to the reader.

T_{14}. $a = (b \Delta b') \Delta a'$.[3]

T_{15}. $a = a' \Delta (b \Delta b')$.

T_{16}. $a \Delta (b \Delta c) = (c \Delta a') \Delta b'$.

T_{17}. $(a \Delta b) \Delta c = b' \Delta (c' \Delta a)$.

T_{18}. *If $a \Delta x = a \Delta y$, or if $x \Delta a = y \Delta a$, then $x = y$.*

T_{19}. *If $a \Delta b = c$, then $a = b \Delta c$ and $b = c \Delta a$.*

[3] T_{14} may be used in place of P_4 in the postulates for groups and Abelian groups.

5. **Postulates for Abelian groups.** The postulates for Abelian groups follow. In Postulate P_6, supply the clause: *whenever the elements involved and their combinations are in K.* The postulates are:

P_0-P_5.

P_6. $a \Delta b = b \Delta a$.

As theorem from P_0-P_6 we have T_{20} following.

T_{20}. $a + b = b + a$.

Proof: D_2, P_6, D_2.

P_0, T_8-T_{13}, T_{20} make $(K, \Delta, ')$ a (non-trivial) Abelian group.

6. **Additional theorems for Abelian groups.** The following are additional theorems for Abelian groups.

T_{21}. $(a \Delta b)' = a' \Delta b'$.

T_{22}. $a \Delta (b \Delta c) = b' \Delta (c \Delta a') = c' \Delta (a' \Delta b) = (a' \Delta b) \Delta c'$.

T_{23}. $a \Delta (b \Delta c) = [(k \Delta a) \Delta b] \Delta (k \Delta c)$.

T_{24}. $(a \Delta b) \Delta (c \Delta d) = a + b + c + d$.

T_{25}. *$(a \Delta b) \Delta (c \Delta d)$ is invariant under the interchange of any two of the elements a, b, c, d.*

T_{26}. *If $a \Delta b = c$, then $b = a \Delta c$ and $a = c \Delta b$.*

T_{27}. $a \Delta (a \Delta b) = (b \Delta a) \Delta a = b$.

7. **Consistency and necessity.** Postulates P_0-P_6 are obviously consistent, and necessary for (non-trivial) Abelian groups. P_0-P_5 are necessary for (non-trivial) groups.

8. **Independence.** The mutual independence of P_0-P_6 is established by systems $\bar{P}_0$-$\bar{P}_6$ in the table below. In this table, system $\bar{P}_i$ is the independence-system for Postulate P_i. In systems $\bar{P}_1$ and $\bar{P}_2$, the blanks indicate absence of K-elements. The table follows.

SYSTEM	K	$a \Delta b$	a'
$\bar{P}_0$	Null class		
$\bar{P}_1$	0, 1	$\begin{array}{c\|cc} & 0 & 1 \\ \hline 0 & - & - \\ 1 & - & - \end{array}$	$\begin{array}{c\|c} a & a' \\ \hline 0 & 0 \\ 1 & 1 \end{array}$
$\bar{P}_2$	0, 1	$\begin{array}{c\|cc} & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array}$	$\begin{array}{c\|c} a & a' \\ \hline 0 & - \\ 1 & - \end{array}$
$\bar{P}_3$	0, 1	$\begin{array}{c\|cc} & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array}$	$\begin{array}{c\|c} a & a' \\ \hline 0 & 0 \\ 1 & 0 \end{array}$

System	K	$a \Delta b$	a'
$\bar{P}_4$	0, 1	$\begin{array}{c\|cc} & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 0 \end{array}$	$\begin{array}{c\|c} a & a' \\ \hline 0 & 0 \\ 1 & 1 \end{array}$
$\bar{P}_5$	0, 1	$\begin{array}{c\|cc} & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array}$	$\begin{array}{c\|c} a & a' \\ \hline 0 & 1 \\ 1 & 0 \end{array}$
$\bar{P}_6$	0, 1, 2, 3, 4, 5	$\begin{array}{c\|cccccc} & 0 & 1 & 2 & 3 & 4 & 5 \\ \hline 0 & 0 & 2 & 1 & 3 & 4 & 5 \\ 1 & 2 & 1 & 0 & 4 & 5 & 3 \\ 2 & 1 & 0 & 2 & 5 & 3 & 4 \\ 3 & 3 & 5 & 4 & 0 & 1 & 2 \\ 4 & 4 & 3 & 5 & 2 & 0 & 1 \\ 5 & 5 & 4 & 3 & 1 & 2 & 0 \end{array}$	$\begin{array}{c\|c} a & a' \\ \hline 0 & 0 \\ 1 & 2 \\ 2 & 1 \\ 3 & 3 \\ 4 & 4 \\ 5 & 5 \end{array}$

The University of California.

CONTRIBUTIONS TO THE THEORY OF HERMITIAN SERIES

By Einar Hille

Introduction

The present paper is devoted to a study of *Hermitian series in the complex domain.* We shall consider expansions of the form

$$f(z) = \sum_{n=0}^{\infty} f_n h_n(z), \tag{1}$$

where

$$h_n(z) = e^{-\frac{1}{2}z^2} H_n(z) = (-1)^n e^{\frac{1}{2}z^2} \frac{d^n}{dz^n} [e^{-z^2}], \tag{2}$$

z being a complex variable.[1] While such series have been rather thoroughly studied for real values of the variable,[2] no adequate discussion of the complex case seems to exist anywhere in the literature. Various aspects of this problem, such as *the domain of convergence, the presence or absence of singularities on the boundary of this domain, gap-theorems,* and *the relations of Hermitian series to associated Dirichlet and Fourier series,* will be discussed in this paper.

The only phase of the complex theory for which the author has been able to find a discussion in the literature is the *representation problem.* Here G. N. Watson [30] and O. Volk [28, 29] have found essentially equivalent sufficient but far from necessary conditions for the representability of a given analytic function by a Hermitian series. We shall not discuss this problem here. We shall solve this problem in the second paper of this series to appear in the Transactions of the American Mathematical Society.

The first problem that confronts us in the complex theory is the *domain of convergence* of a given Hermitian series. While it is well known to analysts working in this field that *the domain of convergence is a strip,* $-\tau < y < \tau$, no simple proof of this fact appears to be available in the literature. Such a proof is given in Chapter 2 of the present paper.

Any convergence proof must be based upon some estimate of *the asymptotic behavior of the Hermitian functions $h_n(z)$ for large values of n.* Such estimates

Received August 4, 1939; presented to the American Mathematical Society, April 7, 1939.

[1] The definition of a Hermitian polynomial differs from one author to the next. I shall employ the notation of an earlier paper of mine [12] to which the reader is referred for the formal properties of Hermitian functions used in this paper. Numbers in brackets refer to the bibliography at the end of the present paper.

[2] A good summary of the theory of Hermitian series in the real case is to be found in Chapter 4 of the treatise of G. Vitali and G. Sansone [26]. The Tchebycheff-Hermite polynomials of these writers differ from those of the present paper by the factor $(-1)^n$.

were first found by G. N. Watson [30, 31] by rather laborious analysis. More recently, N. Schwid [22] gave a thorough investigation of the problem using the methods of R. E. Langer. I prefer to make a new attack on the problem along different lines. I shall use what is essentially a variant of the method of asymptotic integration which I developed in a number of papers many years ago.[3] The point of departure is the differential equation

$$w'' + (2n + 1 - z^2)w = 0 \tag{3}$$

which is satisfied by $h_n(z)$. From this differential equation one concludes that $h_n(z)$ also satisfies a Volterra integral equation

$$h_n(z) = c_n(z) + (2n + 1)^{-\frac{1}{2}} \int_0^z t^2 \sin [(2n + 1)^{\frac{1}{2}}(z - t)]\, h_n(t)\, dt, \tag{4}$$

where

$$c_n(z) = A_n \cos [(2n + 1)^{\frac{1}{2}}z - \tfrac{1}{2}n\pi] \tag{5}$$

and A_n is a positive constant determined by the initial values of $h_n(z)$ and $h_n'(z)$ at the origin. The Volterra equation can be integrated by the method of successive approximations leading to a rapidly convergent series of which $c_n(z)$ is the first term. For values of z which are sufficiently small in comparison with n this first term also dominates the series. This study of the basic integral equation and its consequences for the asymptotic behavior of $h_n(z)$ is carried out in Chapter 1.

While this method of attack is far from being as powerful as that of Langer-Schwid, it is simple and adequate for our purposes. In addition it has one definite advantage; it emphasizes that the passage from $c_n(z)$ to $h_n(z)$ is a linear functional transformation the analytical properties of which can be studied in detail thanks to the explicit expression furnished by the theory of integral equations. This is of fundamental importance for the theory of Hermitian series.

Equation (3) implies that

$$\delta_z h_n(z) = (2n + 1)h_n(z), \tag{6}$$

where the differential operator

$$\delta_z = z^2 - \frac{d^2}{dz^2}. \tag{7}$$

In other words, the functions $h_n(z)$ are characteristic functions of the operator δ_z, corresponding to the characteristic values $2n + 1$. This is of course well known from the modern quantum theory where $\delta_z = p^2 + q^2$ is the operator of the harmonic oscillator. In Chapter 3 of this paper we study analytic func-

[3] See, for example, Hille [11]. The discussion of asymptotic behavior of the Hermite-Weber functions given in this paper refers to large values of z, however. See also the Addendum at the end of this paper.

tions of this operator. *If $G(w)$ is an entire function, then a necessary and sufficient condition that $G(\delta_z)\cdot f(z)$ shall exist and be holomorphic wherever $f(z)$ is holomorphic, and this for all analytic functions $f(z)$, is that $G(w)$ be of order $\leq \frac{1}{2}$ and of minimal type if its order equals $\frac{1}{2}$.* These differential operators $G(\delta_z)$ play a similar rôle in the theory of Hermitian series as the operators $H(d/dz)$ in the theory of power series and Dirichlet series where $H(w)$ is an entire function of order one and minimal type.[4] That in the former case the order of the entire function is restricted to $\frac{1}{2}$, whereas it is one in the latter case, is perhaps explained by the fact that the order of the operator δ_z is two, whereas that of d/dz is one. Thus the product of the orders is the same in both cases. If $f(z)$ is restricted to be an entire function of specified order ρ, the maximal order of $G(w)$ must be less than or equal to the conjugate exponent σ determined by the equation

$$\frac{1}{\rho} + \frac{1}{2\sigma} = 1,$$

provided $\rho > 2$, otherwise $\sigma = 1$. This is to be compared with the corresponding situation for operators $H(d/dz)$ studied in detail by H. Muggli [15] with the aid of the methods of Pólya [19]. The details of this discussion of $G(\delta_z)f(z)$ for entire functions $f(z)$ will be published elsewhere as it has very little immediate bearing on the main problem of the present paper.

The results of Chapter 3 are utilized in Chapter 4 for a study of special types of *regularity preserving factor sequence transformations for Hermitian series.* Given a sequence $\{a_n\}$ such that $\lim n^{-\frac{1}{2}} \log |a_n| = 0$, we form the transformed series

$$F(z) = \sum_{n=0}^{\infty} a_n f_n h_n(z). \tag{8}$$

This series has the same strip of convergence as series (1) thanks to the choice of the a's. What additional conditions must be imposed on the a's in order that $F(z)$ shall have the same singularities as $f(z)$ or at least no other singularities than those of $f(z)$? In other words we ask when it is possible to perform the analytic continuation of $F(z)$ along any finite path along which $f(z)$ can be so continued. The problem is a familiar one in the theory of power and Dirichlet series and the results which we arrive at are analogous to well-known theorems due to Leau, Faber, Fabry, Cramér, Pólya, Ostrowski and others.

In Chapter 5 these results are utilized in order to prove the *existence of singularities of Hermitian series on the boundary of the strip of convergence* under various assumptions on the coefficients. We obtain analogues of well-known theorems by Pringsheim, Vitali, Dienes and others. Finally, we prove several

[4] Though the formal theory of differential operators of infinite order is quite old, rigorous applications to the analytic continuation problem appear for the first time in H. Cramér [3]. This field has been dominated during the last fifteen years by the investigations of G. Pólya and his pupils.

gap theorems of which the following analogue of the theorem of Fabry should be mentioned. *The series*

$$f(z) = \sum_{k=1}^{\infty} f_{n_k} h_{n_k}(z) \tag{9}$$

has its strip of convergence as natural domain of existence if

$$\frac{n_k}{k^2} \to \infty \quad \textit{and} \quad \liminf \, [(n_{k+1})^{\frac{1}{2}} - (n_k)^{\frac{1}{2}}] > 0. \tag{10}$$

In Chapter 6 we return once more to our starting point, the basic integral equation. The relation between $h_n(z)$ and $c_n(z)$ suggests the introduction of *the associated Fourier series*

$$C(z) = \sum_{n=0}^{\infty} f_n c_n(z), \qquad S(z) = \sum_{n=0}^{\infty} f_n s_n(z), \tag{11}$$

where $s_n(z)$ is obtained by replacing cos by sin in formula (5). We also introduce *the associated Dirichlet series*

$$E^+(z) = C(z) + iS(z), \qquad E^-(z) = C(z) - iS(z). \tag{12}$$

We obtain explicit representations of the linear transformations joining $f(z)$ with its associated functions. From these representations we obtain in particular that $f(z)$ *is holomorphic in the cross section of the principal Mittag-Leffler stars of* $C(z)$ *and* $C(-z)$. Further, if z_0 and $-z_0$ are a pair of nearest vertices in the cross section star, then at least one of these points is a singularity of $f(z)$. From this result we conclude, for instance, that conditions (10) of the gap theorem are the best of their kind. Further, *there exist entire functions representable by Hermitian series having a strip of convergence of finite width*, i.e., there need not be any singular points on or anywhere near the lines of convergence.

A number of problems have presented themselves in connection with this investigation which had to be omitted for one reason or another. *The representation problem, Hadamard's multiplication theorem, the relations between coefficients and singularities, overconvergence and gap theorems, the extensions to Laguerre series and to Hermite-Stieltjes integrals,* further study of *the differential operators*, all these problems call for special investigations, some of which are already in progress. The author hopes to return to these problems shortly.

Chapter 1. The asymptotic behavior of $h_n(z)$ for large n

1.1. **The basic integral equation.** It is well known that the function

$$h_n(z) = (-1)^n e^{\frac{1}{2}z^2} \frac{d^n}{dz^n} [e^{-z^2}] \tag{1.1.1}$$

satisfies the differential equation

$$w'' + (2n + 1 - z^2)w = 0. \tag{1.1.2}$$

The positive square root of $2n + 1$ is going to occur throughout this paper and will be denoted by ν_n or ν whenever such an abbreviation can be used without ambiguity.

Let $w_0(z)$ be a given solution of the equation

$$w'' + \nu^2 w = 0 \tag{1.1.3}$$

and form the integral equation

$$w(z) = w_0(z) + \frac{1}{\nu}\int_{z_0}^{z} t^2 \sin \nu(z - t)\, w(t)\, dt. \tag{1.1.4}$$

It is easily seen that the solution of (1.1.4) is uniquely determined and coincides with that particular solution of (1.1.2) which is defined by the initial conditions

$$w(z_0) = w_0(z_0), \qquad w'(z_0) = w_0'(z_0). \tag{1.1.5}$$

Here z_0 is an arbitrary but fixed complex number.

The method of successive approximations shows that the solution of (1.1.4) can be written in the form

$$w(z) = \sum_{k=0}^{\infty} w_k(z), \tag{1.1.6}$$

where

$$w_k(z) = \frac{1}{\nu}\int_{z_0}^{z} t^2 \sin \nu(z - t)\, w_{k-1}(t)\, dt \tag{1.1.7}$$

or

$$w_k(z) = \int_{z_0}^{z}\int_{z_0}^{t_1} \cdots \int_{z_0}^{t_{k-1}} (t_1 t_2 \cdots t_k)^2 K(T_k, n)\, w_0(t_k)\, dt_k\, dt_{k-1} \cdots dt_1 \tag{1.1.8}$$

with

$$K(T_k, n) = \nu^{-k} \prod_{j=1}^{k} \sin \nu(t_{j-1} - t_j), \qquad t_0 = z. \tag{1.1.9}$$

It is a familiar fact that such a series is rapidly convergent, and we shall not spend any time on giving a convergence proof for the general case which is of no interest to us. The solution of (1.1.4) is of importance to us from two different points of view, one more formal, the other purely analytical. The formulas (1.1.6) to (1.1.9) can be regarded as defining a linear transformation on the solutions of the equation (1.1.3) to the solutions of (1.1.2). This trivial observation has important consequences for the theory of Hermitian series which will be presented in Chapter 6 of the present paper. Secondly, the series is asymptotic in nature within certain ranges of z and n, i.e., the first term dominates the series and still better approximations can be found by using partial sums of low order. This property will be verified in the instances of special interest to the Hermitian series in the subsequent paragraphs.

It should be noted that the asymptotic property does not hold unless z_0 and z are relatively small compared to n. The calculations show that preferably z_0 and z should be $o(n^{\frac{1}{6}})$. If this condition is violated, the approximation furnished by the first term of the series is no longer a good one and the problem of asymptotic representation must be tackled by other means. For such questions we refer the reader to the paper by N. Schwid [22]. For the purposes of the present paper, the results obtained by our method are amply sufficient.

1.2. **The case of $h_n(z)$.** Since

$$H_n(z) = \sum_{2k \leq n} (-1)^k \binom{n}{2k} \frac{(2k)!}{k!} (2z)^{n-2k}, \tag{1.2.1}$$

we see that

$$\begin{aligned} H_{2m}(0) &= (-1)^m \frac{(2m)!}{m!}, \qquad H'_{2m}(0) = 0, \\ H_{2m+1}(0) &= 0, \qquad H'_{2m+1}(0) = (-1)^m 2 \frac{(2m+1)!}{m!}. \end{aligned} \tag{1.2.2}$$

Let us define

$$A_{2m} = | H_{2m}(0) |, \qquad A_{2m+1} = | H'_{2m+1}(0) | (4m+3)^{-\frac{1}{2}}, \tag{1.2.3}$$

$$c_n(z) = A_n \cos [(2n+1)^{\frac{1}{2}} z - \tfrac{1}{2} n\pi], \tag{1.2.4}$$

$$s_n(z) = A_n \sin [(2n+1)^{\frac{1}{2}} z - \tfrac{1}{2} n\pi]. \tag{1.2.5}$$

It is clear that both $c_n(z)$ and $s_n(z)$ are solutions of (1.1.3). Further

$$h_n(0) = c_n(0), \qquad h'_n(0) = c'_n(0).$$

Hence, if we choose $z_0 = 0$, $w_0(z) = c_n(z)$, then the corresponding solution of (1.1.4) is $h_n(z)$. It follows that $h_n(z)$ admits of an expansion of the form (1.1.6). The functions $w_{n,k}(z)$ entering in this expansion have a fairly simple structure. Putting

$$w_{n,k}(z) = \nu^{-4k} \omega_{n,k}(\nu z), \tag{1.2.6}$$

we find that

$$\omega_{n,k}(u) = \int_0^u v^2 \sin (u - v)\, \omega_{n,k-1}(v)\, dv, \qquad \omega_{n,0}(u) = A_n \cos (u - \tfrac{1}{2} n\pi). \tag{1.2.7}$$

It is clear that $\omega_{n,k}(u)$ is going to be a linear combination of $\sin u$ and $\cos u$ with coefficients which are polynomials in u, one of which will be of degree $3k$ and the other of degree $3k - 1$. The coefficients of the polynomials are real, rational numbers and their signs appear to alternate. Thus we get finally

$$h_n(z) = c_n(z) \sum_{k=0}^{\infty} P_k(\nu z) \nu^{-4k} + s_n(z) \sum_{k=0}^{\infty} Q_k(\nu z) \nu^{-4k}. \tag{1.2.8}$$

Here

$$P_0(u) = 1, \qquad Q_0(u) = 0,$$
$$P_1(u) = \tfrac{1}{4}u^2 - 1, \qquad Q_1(u) = \tfrac{1}{8}u^3 - \tfrac{1}{4}u,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$P_k(u)$ is an even and $Q_k(u)$ an odd polynomial, one of degree $3k$, the other of degree $3k - 1$, according to the parity of these numbers.

Formulas of this type seem to be new in the theory of Hermitian polynomials though they should be compared with the asymptotic formulas of Adamoff [1] and, in particular, those of Plancherel-Rotach [18]. It should be noted, however, that (1.2.8) is a convergent expansion and not of the asymptotic semi-convergent type familiar in the theory of Bessel functions.

We shall discuss the cases in which z is either real or purely imaginary in more detail. Suppose first that z is real, $z = x$. A simple calculation based on (1.2.7) shows that

$$(1.2.9) \qquad \left| h_n(x) - \sum_{k=0}^{m} \nu^{-4k} \omega_{n,k}(\nu x) \right| \leqq A_n \sum_{m+1}^{\infty} \frac{|x|^{3k}}{(3\nu)^k k!}.$$

In particular,

$$(1.2.10) \qquad | h_n(x) - c_n(x) | \leqq A_n \left\{ \exp\left(\frac{|x|^3}{3\nu} \right) - 1 \right\}.$$

1.3. **The behavior of $h_n(iy)$.** Let us define

$$(1.3.1) \qquad \gamma_n(y) = \tfrac{1}{2}A_n[e^{\nu y} + (-1)^n e^{-\nu y}].$$

We shall prove

THEOREM 1.1. *For positive values of y*

$$(1.3.2) \qquad 0 < i^{-n} h_n(iy) - \gamma_n(y) < A_n e^{\nu y}\left[\exp\left(\frac{y^3}{6\nu} \right) - 1 \right].$$

Proof. Formula (1.2.1) shows that $i^{-n}h_n(iy) > 0$ for $y > 0$. Put $h_n(iy) = i^n \eta_n(y)$. We note that $c_n(iy) = i^n \gamma_n(y)$. Consequently equation (1.1.4) becomes

$$(1.3.3) \qquad \eta_n(y) = \gamma_n(y) + \frac{1}{\nu} \int_0^y u^2 \sinh \nu(y - u)\, \eta_n(u)\, du.$$

All the quantities involved in this equation are positive when $y > 0$. This proves the first part of the required inequality. In order to prove the second half, let us consider formulas (1.2.6) and (1.2.7). Putting $\omega_{n,k}(iu) = \tilde{\omega}_{n,k}(u)$, we get

$$\tilde{\omega}_{n,k}(u) = \int_0^u v^2 \sinh (u - v)\, \tilde{\omega}_{n,k-1}(v)\, dv,$$

and by induction

$$0 < \bar{\omega}_{n,k}(u) < A_n e^u \frac{u^{3k}}{6^k k!}. \tag{1.3.4}$$

The second half of the inequality follows from this estimate.

For values of y which are large in comparison with n, Theorem 1.1 gives a rather poor upper bound for $i^{-n}h_n(iy)$. But by formula (1.2.1)

$$i^{-n}h_n(iy) < i^{-n}h_n(i)y^n e^{\frac{1}{2}\nu^2}, \qquad y > 1, \tag{1.3.5}$$

and $i^{-n}h_n(i)$ can be estimated with the aid of Theorem 1.1.

The series in (1.2.8) can obviously be differentiated term by term. Since

$$\bar{\omega}'_{n,k}(u) = \int_0^u v^2 \cosh(u - v)\, \bar{\omega}_{n,k-1}(v)\, dv,$$

we find by induction that

$$0 < \bar{\omega}'_{n,k}(u) < 2A_n e^u \frac{u^{3k}}{6^k k!}. \tag{1.3.6}$$

Putting

$$\sigma_n(y) = \tfrac{1}{2}A_n[e^{\nu y} - (-1)^n e^{-\nu y}], \tag{1.3.7}$$

and utilizing formula (1.3.6), we obtain

THEOREM 1.2. *For positive values of* y

$$0 < i^{-n+1}h'_n(iy) - \nu\sigma_n(y) < 2\nu A_n e^{\nu y}\left[\exp\left(\frac{y^3}{6\nu}\right) - 1\right]. \tag{1.3.8}$$

Combining these two theorems, we get

THEOREM 1.3. *For* $0 < \epsilon \leq y \leq 1/\epsilon$

$$\frac{h'_n(iy)}{\nu h_n(iy)} = -i + O\left(\frac{1}{\nu}\right) \qquad as \quad n \to \infty, \tag{1.3.9}$$

uniformly in y. *The relation is also true if* i *is replaced by* $-i$.

1.4. **The behavior of $h_n(z)$ on horizontal lines.** We shall study $h_n(x + iy)$ for y fixed, $y = y_0 > 0$. We use (1.1.4) with $z_0 = iy_0$ and

$$w_0(z) = h_n(iy_0) \cos \nu(z - iy_0) + \frac{1}{\nu} h'_n(iy_0) \sin \nu(z - iy_0).$$

The maximum of the absolute value of this expression for $y = y_0$ equals the larger of the two quantities $|h_n(iy_0)|$ and $|h'_n(iy_0)|/\nu$, the ratio of which tends to 1 by Theorem 1.3.

Substituting these values of z_0 and $w_0(z)$ in (1.1.4), we get $h_n(z) = \sum w_k(z)$, where now

$$w_k(x + iy_0) = \frac{1}{\nu}\int_0^x (t + iy_0)^2 \sin \nu(x - t)\, w_{k-1}(t + iy_0)\, dt,$$

whence by complete induction

$$| w_k(x + iy_0) | \leqq \frac{1}{\nu^k k!}[\tfrac{1}{3}|x|^3 + |x|\, y_0^2]^k \cdot \max_t | w_0(t + iy_0) |.$$

Consequently

$$(1.4.1)\quad \begin{aligned} &\left| h_n(x + iy_0) - h_n(iy_0)\cos \nu x - \frac{1}{\nu} h_n'(iy_0)\sin \nu x \right| \\ &\leqq \max\left[| h_n(iy_0) |, \frac{1}{\nu} | h_n'(iy_0) | \right]\cdot\left\{\exp\left[\frac{1}{\nu}(\tfrac{1}{3}|x|^3 + |x|\, y_0^2)\right] - 1\right\}. \end{aligned}$$

Using Theorem 1.3 we see that

$$h_n(iy_0)\cos \nu x + \frac{1}{\nu} h_n'(iy_0)\sin \nu x = h_n(iy_0)\left\{e^{-i\nu x} + O\left(\frac{1}{\nu}\right)\right\}.$$

From these estimates we get

THEOREM 1.4. *For $y \neq 0$*

$$(1.4.2)\qquad e^{\pm i\nu x}\frac{h_n(x + iy)}{h_n(iy)} = 1 + O\left(\frac{1}{\nu}\right) \qquad \text{as} \quad n \to \infty,$$

the sign in the exponent being the same as that of y. The relation holds uniformly with respect to x and y in the regions $-1/\epsilon \leqq x \leqq 1/\epsilon$, $0 < \epsilon \leqq | y | \leqq 1/\epsilon$.

Theorem 1.1 implies that

$$h_n(iy) = c_n(iy)\left[1 + O\left(\frac{1}{\nu}\right)\right] \qquad \text{as } n \to \infty.$$

Since

$$e^{\mp i\nu x} c_n(iy) = c_n(x + iy)\{1 + O(e^{-\nu|y|})\},$$

we can restate Theorem 1.4 as follows:

THEOREM 1.5. *For $z = x + iy$, $y \neq 0$,*

$$(1.4.3)\qquad h_n(z) = c_n(z)\left[1 + O\left(\frac{1}{\nu}\right)\right] \qquad \text{as } n \to \infty.$$

The relation holds uniformly with respect to z in any bounded closed region having no point in common with the real axis. For real values of z, $z = x$, (1.4.3) is to be replaced by

$$(1.4.4)\qquad h_n(x) = c_n(x) + O\left(\frac{A_n}{\nu}\right).$$

If a higher degree of approximation is desired than is furnished by this theorem, we can obtain the desired result from formula (1.2.8). The estimate of the remainder can be obtained with the aid of (1.2.7) in the usual manner and the result is

THEOREM 1.6. *For all values of* z

$$\left| h_n(z) - c_n(z) \sum_{k=0}^{m} P_k(\nu z)\nu^{-4k} - s_n(z) \sum_{k=0}^{m} Q_k(\nu z)\nu^{-4k} \right| \leq A_n e^{\nu|y|} \sum_{k=m+1}^{\infty} \frac{|z|^{3k}}{(3\nu)^k k!}. \tag{1.4.5}$$

Chapter 2. Convergence theory

2.1. The strip of convergence. Let

$$\sum_{n=0}^{\infty} f_n h_n(z) \tag{2.1.1}$$

be a given Hermitian series. We shall determine the domain of convergence of this series with the aid of the asymptotic relations of the preceding chapter.

There is a classical theorem in the theory of power series according to which the boundedness of the terms at a point $z_0 \neq 0$ implies the absolute convergence of the series in the circle $|z| < |z_0|$. This theorem has the following analogue in the theory of Hermitian series.

THEOREM 2.1. *If* $|f_n h_n(z_0)| \leq M$ *for all* n *and* $z_0 = x_0 + iy_0$, *where* $y_0 \neq 0$, *then the series* (2.1.1) *is absolutely convergent in the strip* $-|y_0| < y < |y_0|$.

Proof. The theorem is a simple consequence of Theorem 1.5 which says that we can find a positive constant a such that for all large n

$$|h_n(x_0 + iy_0)| > a|c_n(x_0 + iy_0)| > \tfrac{1}{2}aA_n \exp[\nu|y_0|].$$

Similarly if $z = x + iy$, $y \neq 0$, we can find a positive constant A such that for all large n

$$|h_n(x + iy)| < A|c_n(x + iy)| < AA_n \exp[\nu|y|].$$

When $y = 0$ we still have an estimate of the form $|h_n(x)| < AA_n$. Consequently,

$$|f_n h_n(x + iy)| < \frac{2}{a} AM \exp[-\nu(|y_0| - |y|)].$$

This estimate proves the convergence of (2.1.1) for $-|y_0| < y < |y_0|$.

From this theorem it follows that the set of convergence of a Hermitian series is a strip $-\tau < y < \tau$ with the possible addition of points on the lines $y = \pm\tau$. We shall call τ the *ordinate of convergence* and the lines $y = \pm\tau$ the *lines of convergence* of the given series. It remains to determine τ. This is done in

THEOREM 2.2.

$$\tau = -\limsup_{n\to\infty} \frac{1}{\nu_n} \log [A_n |f_n|]. \tag{2.1.2}$$

Remark. We can clearly replace $\nu_n = (2n+1)^{\frac{1}{2}}$ by $(2n)^{\frac{1}{2}}$ and instead of A_n we can write $2^n\Gamma(\frac{1}{2}n + a)$, where a is an arbitrary fixed real quantity, or simply $(2n/e)^{\frac{1}{2}n}$. We shall use these equivalent forms whenever they are preferable.

We base the proof of this theorem upon the following

LEMMA. *Let $\{\lambda_n\}$ be a monotone increasing sequence of positive numbers, $0 < \lambda_n < \lambda_{n+1}$, such that $\lambda_n/\log n \to \infty$. For the given series $\sum u_n$, $u_n \geqq 0$, form*

$$L = \limsup_{n\to\infty} \frac{1}{\lambda_n} \log u_n .$$

If $L < 0$ the series $\sum u_n$ converges, if $L > 0$ it diverges, and there is indetermination when $L = 0$.

The classical case $\lambda_n = n$ is well known, and the general case is easily proved by noticing that the assumptions on λ_n imply that $\sum e^{\lambda_n x}$ converges for $x < 0$ and diverges for $x > 0$.

Proof of Theorem 2.2. We apply the lemma with $\lambda_n = \nu_n$ to the case $u_n = |f_n h_n(z)|$ with $z = x + iy$, $y \neq 0$. We have

$$\frac{1}{\nu} \log |f_n h_n(z)| = \frac{1}{\nu} \log [A_n |f_n|] + \frac{1}{\nu} \log \frac{|h_n(iy)|}{A_n} + \frac{1}{\nu} \log \left| \frac{h_n(z)}{h_n(iy)} \right| .$$

By Theorem 1.1 the second term on the right tends to $|y|$ as $n \to \infty$, whereas the last term tends to zero by Theorem 1.4. Hence

$$\limsup_{n\to\infty} \frac{1}{\nu_n} \log |f_n h_n(z)| = -\tau + |y|.$$

Thus we have convergence for $|y| < \tau$. That we have divergence for $|y| > \tau$ follows from the fact that there will then be infinitely many terms in the series of absolute value greater than one. Thus τ is the ordinate of convergence of the series. The letter τ will always be used in this sense in the following.

THEOREM 2.3. *If a Hermitian series is absolutely convergent at any one point on one of the lines of convergence, then it converges absolutely at all points of the lines of convergence.*

This is an immediate consequence of Theorem 1.4.

2.2. **An equiconvergence theorem.** The discussion in the preceding section is elementary and it has the advantage of paralleling the familiar discussion of power series. It is unnecessarily long, however, and the main theorem could be derived more directly. It is an immediate consequence of the following

THEOREM 2.4. *The series* (2.1.1) *converges for a non-real value of z if and only if the associated series*

$$\sum_{n=0}^{\infty} f_n c_n(z) \equiv C(z) \tag{2.2.1}$$

converges for this value.

Proof. (i) *Absolute convergence.* If one series is absolutely convergent so is the other by virtue of Theorem 1.5.

It is easy to obtain the theorems of §2.1 from this observation. We can write $C(z)$ as the sum of two Dirichlet series

$$\tfrac{1}{2} \sum_{n=0}^{\infty} i^{-n} f_n A_n e^{i\nu z} \quad \text{and} \quad \tfrac{1}{2} \sum_{n=0}^{\infty} i^{n} f_n A_n e^{-i\nu z} \tag{2.2.2}$$

of which the former converges in a half plane $y > \alpha$ and the latter in $y < \beta$. If $\lambda_n/\log n \to \infty$, the abscissa of absolute convergence of the Dirichlet series

$$\sum a_n e^{-\lambda_n s}$$

coincides with the abscissa of ordinary convergence and is given by

$$\sigma = \limsup_{n\to\infty} \frac{1}{\lambda_n} \log |a_n|.$$

Applying this formula to the present case, we find that $\alpha = -\tau$ and $\beta = \tau$. Hence the series (2.2.1) is absolutely convergent for $-\tau < y < \tau$ and diverges outside of this strip. The same then is true of the Hermitian series. This is an alternative proof of Theorems 2.1 and 2.2.

(ii) *Non-absolute convergence.* Suppose that (2.2.1) converges at a point z_0 on the boundary of the strip of convergence without being absolutely convergent. Now if $\sum a_n$ is a convergent series and if $\{\epsilon_n\}$ is a monotone decreasing sequence of positive numbers, then $\sum \epsilon_n a_n$ is also convergent. Hence the convergence of

$$\sum_{n=0}^{\infty} f_n c_n(z_0) \quad \text{implies that of} \quad \sum_{n=0}^{\infty} \nu_n^{-k} f_n c_n(z_0)$$

for every $k \geqq 0$. Further the Dirichlet series (2.2.2) must converge for $z = z_0$ and hence also

$$\sum_{n=0}^{\infty} \nu_n^{-k} f_n s_n(z_0)$$

for every $k \geqq 0$.

We now appeal to Theorem 1.6. Taking $m = 2$ we get

$$h_n(z) = c_n(z)\{1 + P_1(\nu z)\nu^{-4} + P_2(\nu z)\nu^{-8}\}$$
$$+ s_n(z)\{Q_1(\nu z)\nu^{-4} + Q_2(\nu z)\nu^{-8}\} + R_{n,3}(z),$$

where

$$| R_{n,3}(z) | \leqq M(z) A_n \nu^{-3} e^{\tau |y|},$$

$M(z)$ being independent of n. Here we can replace $s_n(z)$ by $\pm i c_n(z)$, the sign being opposite that of y, making an error which will be $O\{A_n e^{-\tau|y|}\}$. Further, in the polynomials we can suppress the constant term of P_1, the first degree term of Q_1, and all terms of degree $\leqq 5$ in P_2 and Q_2 without affecting the order of magnitude of the remainder. Hence

$$(2.2.3) \quad h_n(z) = c_n(z)\{1 + (a_0 + a_3 z^3)\nu^{-1} + (a_2 z^2 + a_6 z^6)\nu^{-2}\} + R_n(z),$$

where the a's are numerical constants and

$$(2.2.4) \qquad | R_n(z) | \leqq M(z) A_n \nu^{-3} e^{\tau|y|}.$$

Multiplying by f_n and adding we get

$$(2.2.5) \quad \begin{aligned} \sum_0^N f_n h_n(z) &= \sum_0^N f_n c_n(z) + (a_0 + a_3 z^3) \sum_0^N \nu^{-1} f_n c_n(z) \\ &\quad + (a_2 z^2 + a_6 z^6) \sum_0^N \nu^{-2} f_n c_n(z) + \sum_0^N f_n R_n(z). \end{aligned}$$

As $N \to \infty$ the first three terms on the right tend to finite limits for $z = z_0$ and (2.2.4) shows that this is also the case of the last sum. Hence the convergence of (2.2.1) implies that of (2.1.1).

Suppose conversely that (2.1.1) converges for a non-real value $z = z_0$. We can still use (2.2.5) and conclude that the last sum tends to a finite limit for $z = z_0$ as $N \to \infty$. But the convergence of $\sum f_n h_n(z)$ implies that of $\sum \nu^{-k} f_n h_n(z)$ for every $k \geqq 0$. A simple calculation shows that

$$\begin{aligned} &\sum_0^N f_n h_n(z) - (a_0 + a_3 z^3) \sum_0^N \nu^{-1} f_n h_n(z) \\ &\quad + (a_0 - a_2 z^2 + a_3 z^3 - a_6 z^6) \sum_0^N \nu^{-2} f_n h_n(z) = \sum_0^N f_n c_n(z) + \sum_0^N f_n R_n^*(z), \end{aligned}$$

where $R_n^*(z)$ also satisfies (2.2.4). As $N \to \infty$, the sums on the left and the second sum on the right tend to finite limits for $z = z_0$. Hence the convergence of (2.1.1) implies the convergence of (2.2.1).

Thus we see that the discussion of the convergence of a Hermitian series on the lines of convergence bordering its strip of convergence is reduced to the apparently much simpler discussion of the associated Fourier series (2.2.1) or the associated Dirichlet series (2.2.2). The reader should note that our results do not apply if the set of convergence of the Hermitian series reduces to a point set on the real axis. For this case, however, G. Szegö has proved an equiconvergence theorem to which we have nothing to add.[5] Theorem 2.4 can be regarded as a complex analogue of this theorem.

[5] G. Szegö [24], pp. 114–115.

Chapter 3. A class of differential operators

3.1. **The differential operator δ_z.** We have seen that $h_n(z)$ satisfies the differential equation (1.1.2). Let us introduce the differential operator

$$\delta_z = z^2 - \frac{d^2}{dz^2}. \tag{3.1.1}$$

We have

$$\delta_z h_n(z) = (2n + 1)h_n(z). \tag{3.1.2}$$

In other words, $h_n(z)$ is a *characteristic function* of the operator δ_z corresponding to the *characteristic value* $2n + 1$.

Thus if

$$f(z) = \sum_{n=0}^{\infty} f_n h_n(z)$$

is a given Hermitian series whose ordinate of convergence is $\tau > 0$, we have clearly the right to apply the operator δ_z termwise obtaining

$$\delta_z f(z) = \sum_{n=0}^{\infty} (2n + 1) f_n h_n(z),$$

valid within the strip of convergence of the original series. *This operator consequently has the effect of multiplying the n-th coefficient of a Hermitian series by* $2n + 1$.

The k-th power of δ_z will be denoted by δ_z^k. We have clearly

$$\delta_z^m \cdot \delta_z^n = \delta_z^n \cdot \delta_z^m = \delta_z^{m+n}.$$

Let $P(w)$ be a polynomial with constant coefficients, $P(w) = \sum_{j=0}^{k} a_j w^j$, and form the operator

$$P(\delta_z) = \sum_{j=0}^{k} a_j \delta_z^j, \tag{3.1.3}$$

where δ_z^0 is the identity. It is an easy matter to show that

$$P(\delta_z) f(z) = \sum_{n=0}^{\infty} P(2n + 1) f_n h_n(z),$$

again valid within the strip of convergence of the series for $f(z)$. We shall see later that corresponding formulas hold for certain classes of entire functions of δ_z.

The present chapter will be devoted to a study of those properties of δ_z and of functions of δ_z which will be useful to us in the study of singularities of functions defined by Hermitian series.

The basic property of δ_z in this connection is that it preserves holomorphism in the finite part of the plane. In other words, if the analytic function $f(z)$ is holomorphic, i.e., single-valued, continuous, and differentiable, in the domain D,

not containing $z = \infty$, then $\delta_z f(z)$ is an analytic function holomorphic in D. The theory of linear differential equations shows that we have the following converse of this statement. If $g(z)$ is an analytic function, holomorphic in a simply-connected domain Δ not containing $z = \infty$, and if $z_0 \in \Delta$ and a and b are given complex numbers, then there exists one and only one analytic function $f(z)$, holomorphic in Δ, such that $\delta_z f(z) = g(z), f(z_0) = a, f'(z_0) = b$.

These properties extend to polynomials in δ_z with constant coefficients. Let $P(\delta_z)$ be defined by (3.1.3) and let $f(z)$ be holomorphic in the circle $| z - z_0 | < R$. Then $P(\delta_z)f(z) \equiv g(z)$ is holomorphic in the same circle. If $f(z)$ can be continued analytically along a bounded Jordan arc L from $z = z_0$ to $z = z_1$, then $g(z)$ can be continued along the same arc. On the other hand, if $z = z_1$ is a singular point of this particular branch of $f(z)$ for continuation along L, then the same is true of $g(z)$, and z_1 cannot be a singular point of $g(z)$ unless it is singular for $f(z)$. In this sense the operator neither introduces nor removes singular points. It clearly preserves the property of being single-valued, but it may very well transform a multiple-valued function into a single-valued one.

We begin our discussion with an elementary question, that of expressing $f(z)$ in terms of $g(z)$ in the case in which $P(w) = w^k$. In the remainder of the chapter we are concerned with the question of extending some of the results for polynomial operators to more general classes. Here convergence questions become of importance and the first problem we have to solve is to estimate as accurately as possible the rate of growth of $\delta_z^k f(z)$ with respect to k under different assumptions on $f(z)$. We then determine certain classes of entire functions $G(w)$ such that for every analytic function $f(z)$ the transformed function $G(\delta_z)f(z)$ is holomorphic wherever $f(z)$ is. The results referring to the case in which $f(z)$ is an entire function of given order are only indicated as they are of no importance for the singularity problem and the author intends to treat the questions involved in full in another publication.

3.2. **The equation $\delta_z^k F(z) = f(z)$.** We shall derive explicit formulas for the solution $F(z)$ of the differential equation

(3.2.1) $$\delta_z^k F(z) = f(z),$$

where $f(z)$ is a given function.

We start with the case $k = 1$, i.e.,

(3.2.2) $$w'' - z^2 w = -f(z).$$

It is well known that the corresponding homogeneous differential equation

(3.2.3) $$w'' - z^2 w = 0$$

is satisfied by

(3.2.4) $$w_1(z) = z^{\frac{1}{2}} J_{-\frac{1}{4}}(\tfrac{1}{2}z^2), \qquad w_2(z) = z^{\frac{1}{2}} J_{\frac{1}{4}}(\tfrac{1}{2}z^2),$$

where $J_\alpha(u)$ is the ordinary Bessel function of order α. These functions form a fundamental system.

Let us introduce the function

$$G(z, t) = \frac{\pi\sqrt{2}}{4}\{w_1(z)w_2(t) - w_1(t)w_2(z)\}. \tag{3.2.5}$$

A simple calculation shows that

$$\delta_z G(z, t) = \delta_t G(z, t) = 0, \qquad \left\{\frac{\partial}{\partial z} G(z, t)\right\}_{t=z} = -1. \tag{3.2.6}$$

Suppose now that $f(z)$ is a given analytic function, holomorphic in a bounded simply-connected domain Δ, and let $z_0 \in \Delta$. We shall determine that solution $F(z)$ of (3.2.2) for which $F(z_0)$ and $F'(z_0)$ take on preassigned values. Let $W_1(z)$ be that solution of (3.2.3) for which

$$W_1(z_0) = F(z_0), \qquad W_1'(z_0) = F'(z_0). \tag{3.2.7}$$

A simple calculation based upon (3.2.6) shows that

$$F(z) = W_1(z) + \int_{z_0}^{z} G(z, t)f(t)\, dt \tag{3.2.8}$$

is the required solution of (3.2.2). This function is holomorphic in Δ and can be continued analytically along any path L from $z = z_0$ along which $f(z)$ can be continued. This is an immediate consequence of the fact that both $W_1(z)$ and $G(z, t)$ are entire functions of z and that the latter is also an entire function of t.

Using (3.2.8) we can now derive the general solution of (3.2.1) by complete induction. We have merely to put

$$\delta_z^{k-1} F(z) = \varphi_1(z)$$

and solve the resulting equation for $\varphi_1(z)$ with the aid of (3.2.8). We then put

$$\delta_z^{k-2} F(z) = \varphi_2(z), \qquad \delta_z \varphi_2(z) = \varphi_1(z),$$

and apply the same process once more, etc. The final result is an expression of the form

$$\begin{aligned} F(z) = W_1(z) &+ \int_{z_0}^{z} G_1(z, t)W_2(t)\, dt + \int_{z_0}^{z} G_2(z, t)W_3(t)\, dt \\ &+ \cdots + \int_{z_0}^{z} G_{k-1}(z, t)W_k(t)\, dt + \int_{z_0}^{z} G_k(z, t)f(t)\, dt. \end{aligned} \tag{3.2.9}$$

Here

$$G_1(z, t) = G(z, t), \qquad G_m(z, t) = \int_t^z G_1(z, u)G_{m-1}(u, t)\, du, \tag{3.2.10}$$

and the $W_m(z)$ are solutions of (3.2.3) determined by the initial conditions

$$\begin{aligned} [\delta_z^{m-1} F(z)]_{z=z_0} &= W_m(z_0), \\ \left[\frac{d}{dz}\delta_z^{m-1} F(z)\right]_{z=z_0} &= W_m'(z_0), \end{aligned} \qquad (m = 1, 2, 3, \cdots, k). \tag{3.2.11}$$

The solution is clearly holomorphic in any bounded simply-connected domain Δ in which $f(z)$ is holomorphic. We note that all the iterated kernels $G_k(z, t)$ are entire functions of z and t.

3.3. **Estimates of $\delta_z^k f(z)$.** In the present section we shall derive suitable expressions for $\delta_z^k f(z)$ from which we can obtain estimates of the rate of growth of this function with respect to k under different assumptions on $f(z)$.

We start with the case in which $f(z)$ is holomorphic in the circle $|z - z_0| < R < \infty$, and let $M(r)$ denote the maximum modulus of $f(z)$ on the circle $|z - z_0| = r < R$, i.e.,

$$M(r) = \max_{0 \leq \theta \leq 2\pi} |f(z_0 + re^{i\theta})|,$$

and put $|z_0| = s$.

From the definition of $\delta_z f(z)$ follows that

$$\delta_z f(z_0) = z_0^2 f(z_0) - \frac{1}{\pi i} \int_C \frac{f(t)\,dt}{(t - z_0)^3}, \tag{3.3.1}$$

where C can be taken as the circle $|t - z_0| = p < R$. If in this formula we replace $f(z)$ by $\delta_z f(z)$, we get

$$\begin{aligned} \delta_z^2 f(z_0) = z_0^4 f(z_0) - \frac{z_0^2}{\pi i} \int_C \frac{f(t)\,dt}{(t - z_0)^3} - \frac{1}{\pi i} \int_C \frac{t^2 f(t)\,dt}{(t - z_0)^3} \\ + \left(\frac{1}{\pi i}\right)^2 \int_{C_2} \int_{C_1} \frac{f(t_1)\,dt_1 dt_2}{(t_1 - t_2)^3 (t_2 - z_0)^3}. \end{aligned} \tag{3.3.2}$$

Here $C_1 : |t_1 - t_2| = p_1$, $C_2 : |t_2 - z_0| = p_2$, where $p_1 + p_2 < R$. By complete induction we obtain

$$\delta_z^k f(z_0) = \sum J^{(m)}_{j_1 \cdots j_m}. \tag{3.3.3}$$

Here $0 \leqq m \leqq k$ and

$$\begin{aligned} J^{(0)} &= z_0^{2k} f(z_0), \\ J^{(1)}_j &= z_0^{2(k-1-j)} \frac{-1}{\pi i} \int_C \frac{t^{2j} f(t)\,dt}{(t - z_0)^3} \qquad (0 \leqq j \leqq k - 1), \\ J^{(2)}_{j_1 j_2} &= z_0^{2(k-2-j_1-j_m)} \left(\frac{-1}{\pi i}\right)^2 \int_{C_2} \int_{C_1} \frac{t_1^{2j_1} t_2^{2j_2} f(t_1)\,dt_1 dt_2}{(t_1 - t_2)^3 (t_2 - z_0)^3} \\ &\qquad (0 \leqq j_1 + j_2 \leqq k - 2), \\ &\cdots\cdots\cdots\cdots \\ J^{(m)}_{j_1 \cdots j_m} &= z_0^{2(k-m-j_1-\cdots-j_m)} \left(\frac{-1}{\pi i}\right)^m \\ &\quad \int_{C_m} \cdots \int_{C_2} \int_{C_1} \frac{t_1^{2j_1} t_2^{2j_2} \cdots t_m^{2j_m} f(t_1)\,dt_1 dt_2 \cdots dt_m}{(t_1 - t_2)^3 (t_2 - t_3)^3 \cdots (t_m - z_0)^3} \\ &\qquad (0 \leqq j_1 + j_2 + \cdots + j_m \leqq k - m), \\ &\cdots\cdots\cdots\cdots \end{aligned} \tag{3.3.4}$$

Thus there are $\binom{k}{m}$ m-fold integrals $J^{(m)}_{j_1 \cdots j_m}$ ($m = 1, 2, 3, \cdots, k$). It is understood that the integration is first carried out with respect to t_1, then with respect to t_2, etc. As contours of integration we can use circles $| t_1 - t_2 | = p_{m,1}$, $| t_2 - t_3 | = p_{m,2}, \cdots, | t_m - z_0 | = p_{m,m}$, with the sole proviso that $\sum p_{m,j} < R$.

If $2m \leqq k$ the integrals are in a convenient form for estimating, but for $2m > k$ they should be further reduced. If $2m > k$, then $k - m < \frac{1}{2}k$ and this means that among the m integers $j_1, j_2, \cdots, j_m$ at least $(2m - k)$ numbers are equal to zero. Suppose that they are not all equal to zero but $j_i \neq 0$ for $i = i_1, i_2, \cdots, i_\mu$ and no other values. To simplify the notation, let us put

$$2j_{i_\alpha} = n_\alpha, \quad n_1 + n_2 + \cdots + n_\mu = N_\mu, \quad i_0 = 0, \quad i_{\mu+1} = m \quad \text{if} \quad i_\mu < m,$$

$$2(i_\alpha - i_{\alpha-1}) = d_\alpha, \quad t_{i_\alpha} = u_\alpha .$$

If $i_\mu < m$ we have then

$$(3.3.5) \qquad \begin{aligned} J^{(m)}_{j_1 \cdots j_m} &= (2\pi i)^{-\mu-1} \prod_{\alpha=1}^{\mu+1} (d_\alpha)!\, z_0^{2(k-m)-N_\mu} \\ &\times \iint \cdots \int \frac{u_1^{n_1} u_2^{n_2} \cdots u_\mu^{n_\mu} f(u_1)\, du_1 du_2 \cdots du_{\mu+1}}{(u_1 - u_2)^{d_1+1}(u_2 - u_3)^{d_2+1} \cdots (u_\mu - u_{\mu+1})^{d_\mu+1}(u_{\mu+1} - z_0)^{d_{\mu+1}+1}}. \end{aligned}$$

This formula is also valid if all $j_i = 0$. We have then $\mu = 0$, $d_1 = 2m$, $N_0 = 0$, and there is a single integral. If $i_\mu = m$, the formula must be modified in an obvious manner. In the outside factors $\mu + 1$ is replaced by μ and there are μ integrations instead of $\mu + 1$, so that the differential $du_{\mu+1}$ drops out and the last two factors in the denominator are replaced by $(u_\mu - z_0)^{d_\mu+1}$.

Let us now proceed to an estimate of these integrals. For this purpose we have to dispose of the circles of integration in a suitable manner. The results are somewhat different according as R is finite or infinite, but it is possible to derive estimates valid for both cases. We do this by introducing two parameters p and q, the values of which will be disposed of later. For the time being it is enough to know that $0 < p, q < R$, that they are independent of $j_1, j_2, \cdots, j_m$, but may depend upon k, and that p enters in the integrals with $2m \leqq k$ and q in those with $2m > k$.

We start with the case $2m \leqq k$. Here we take $J^{(m)}_{j_1 \cdots j_m}$ in its unreduced form and choose

$$p_{m,1} = p_{m,2} = \cdots = p_{m,m} = \frac{p}{m}.$$

Using the fact that $| z_0 | = s < s + p$, $| t_\alpha | \leqq s + p$ ($\alpha = 1, 2, \cdots, m$), we get by standard procedure

$$(3.3.6) \qquad | J^{(m)}_{j_1 \cdots j_m} | \leqq 2^m (s + p)^{2(k-m)} \left(\frac{m}{p}\right)^{2m} M(p).$$

Denoting by S_k^1 the sum of all integrals $J^{(m)}$ with $2m \leqq k$ and by S_k^2 those with $2m > k$, we have

(3.3.7)
$$|S_k^1| \leqq M(p) \sum_{2m \leqq k} \binom{k}{m} 2^m (s+p)^{2(k-m)} \left(\frac{m}{p}\right)^{2m}.$$

We can obtain several useful estimates of this sum. First, we may replace $(m/p)^{2m}$ by its largest value $(k/(2p))^k$ and then extend the summation with respect to m up to k instead of to $[\frac{1}{2}k]$. The result is

(3.3.8)
$$|S_k^1| \leqq M(p) \left(\frac{k}{2p}\right)^k [(s+p)^2 + 2]^k.$$

Another possibility is to replace $(m/p)^{2m}$ by $(k/(2p))^{2m}$ and then sum with respect to m up to k. This gives

(3.3.9)
$$|S_k^1| \leqq M(p) \left[(s+p)^2 + 2\left(\frac{k}{2p}\right)^2 \right]^k.$$

Incidentally, since (3.3.6) is valid for all m and not merely for $2m \leqq k$, we can use the second method to estimate $\delta_z^k f(z_0)$ itself. This device gives

(3.3.10)
$$|\delta_z^k f(z_0)| \leqq M(p) \left[(|z_0| + p)^2 + 2\left(\frac{k}{p}\right)^2 \right]^k,$$

an estimate easy to obtain and easy to handle though as a rule not the best possible.

We now proceed to an estimate of S_k^2. Here we use the reduced form of the integrals. We choose as contours of integration the circles

$$|u_\alpha - u_{\alpha+1}| = \frac{d_\alpha q}{2m} \qquad (\alpha = 1, 2, \cdots, \mu \text{ (or } \cdots, \mu, \mu+1)),$$

where $u_{\mu+1} = z_0$ in the first and $u_{\mu+2} = z_0$ in the second case. Customary methods give

$$|J_{j_1 \cdots j_m}^{(m)}| \leqq \left(\frac{2m}{q}\right)^{2m} \prod_\alpha \{(d_\alpha)! \, d_\alpha^{-d_\alpha}\} (s+q)^{2(k-m)} M(q),$$

where the product extends over μ or $\mu + 1$ factors as the case may be.

But for any integer $d \geqq 1$

$$d! \, d^{-d} \leqq (2\pi d)^{\frac{1}{2}} \exp[-d + (12\, d)^{-1}].$$

Further $\sum d_\alpha = 2m$ and $\sum (1/d_\alpha) < \frac{1}{2}(\mu+1)$, since each $d_\alpha \geqq 2$ and they are μ or $\mu + 1$ in number. The maximum of $\prod d_\alpha$ is reached when all the d's are equal. Combining these observations we get the estimate

$$|J_{j_1 \cdots j_m}^{(m)}| \leqq \left(\frac{2m}{eq}\right)^{2m} \left(\frac{4\pi e m}{\mu+1}\right)^{\frac{1}{2}(\mu+1)} (s+q)^{2(k-m)} M(q)$$

for $2m > k$. Here $\epsilon = e^{\frac{1}{12}}$ and we have taken the case in which there are $\mu + 1$ integrals rather than μ. The estimate holds also in the latter case, however, since the factor involving μ is an increasing function of μ within the range considered. We have $\mu \leqq k - m$, but for the reason just mentioned we can always take $\mu = k - m$ in the formula.

Since $4\pi\epsilon < 16$, we obtain

$$|J_{j_1 \cdots j_m}^{(m)}| \leqq \left(\frac{2m}{eq}\right)^{2m} \left(\frac{16m}{k - m + 1}\right)^{\frac{1}{2}(k-m+1)} (s + q)^{2(k-m)} M(q)$$

and

$$|S_k^2| \leqq M(q) \sum_{2m>k} \binom{k}{m} \left(\frac{2m}{eq}\right)^{2m} \left(\frac{16m}{k - m + 1}\right)^{\frac{1}{2}(k-m+1)} (s + q)^{2(k-m)}.$$

Introducing $k - m = j$, we can write this sum as follows

$$2^{2k+2} \sum_{2j \leqq k} \frac{k!}{j!\,(k-j)!} (eq)^{2(j-k)} (k - j)^{2k - \frac{1}{2}(3j-1)} (j + 1)^{-\frac{1}{2}(j+1)} (s + q)^{2j}.$$

But for $1 \leqq j \leqq \frac{1}{2}k$

$$k!/(k - j)! < k^j \quad \text{and} \quad (k - j)^{2k - \frac{1}{2}(3j-1)} < k^{2k-j} j^{-\frac{1}{2}(j-1)}.$$

Hence the sum is dominated by

$$4 \left(\frac{2k}{eq}\right)^{2k} \left\{k^{\frac{1}{2}} + \sum \frac{[eq(s + q)]^{2j}}{j^j j!}\right\},$$

where $1 \leqq j \leqq \frac{1}{2}k$. The expression within the braces is evidently dominated by a suitably chosen exponential function and a simple calculation shows that

$$|S_k^2| \leqq 8M(q) \left(\frac{2k}{eq}\right)^{2k} k^{\frac{1}{2}} \exp [4q(s + q)]. \tag{3.3.11}$$

Let us now consider the case in which R is finite. We can then choose $p = q < R$. Formula (3.3.8) gives the best estimate of S_k^1 and combining with (3.3.11) we get

$$|\delta_z^k f(z_0)| \leqq M(p) \left\{\left(\frac{k}{2p}\right)^k [(s + p)^2 + 2]^k + 8 \left(\frac{2k}{ep}\right)^{2k} k^{\frac{1}{2}} \exp [4p(s + p)]\right\}.$$

From this inequality we conclude the existence of an $M(p, z_0)$, finite and independent of k, such that

$$|\delta_z^k f(z_0)| \leqq M(p, z_0)(2k)!\, p^{-2k}, \tag{3.3.12}$$

where p is any quantity less than R, the distance from z_0 to the nearest singularity of $f(z)$.[6] Consequently

$$\limsup_{k\to\infty} k^{-2} |\delta_z^k f(z_0)|^{1/k} \leqq \left(\frac{2}{ep}\right)^2,$$

[6] Let D be a bounded domain such that $f(z)$ is holomorphic at all points of D, the radius of holomorphism having a positive lower bound in D. Then the proof shows that $M(z_0, p)$ is bounded in D. Further, formula (3.3.13) holds uniformly in D.

and since this holds for every $p < R$, we have finally also

$$\limsup_{k\to\infty} k^{-2} \mid \delta_z^k f(z_0) \mid^{1/k} \leqq \left(\frac{2}{eR}\right)^2. \tag{3.3.13}$$

This is the best estimate of its kind in the sense that in the class of functions holomorphic in the circle $\mid z - z_0 \mid < R$ there are members for which the sign of equality holds in (3.3.13) even if lim sup be replaced by lim. We shall construct such functions.

Let $z_0 = x_0 + iy_0$, where we can suppose $x_0 \geqq 0$, $y_0 \geqq 0$ without restriction. At the start, we exclude the case in which z_0 is real positive, so that either $z_0 = 0$ or $y_0 > 0$. Now form

$$f(z) = \sum_{n=m}^{\infty} (-1)^n \exp\left[-(\alpha - ix_0)(4n + 1)^{\frac{1}{2}}\right] \frac{h_{2n}(z)}{A_{2n}}, \tag{3.3.14}$$

where α is real and greater than y_0 and m will be chosen later. The series converges in the strip $-\alpha < y < \alpha$ so that $f(z)$ is certainly holomorphic in $\mid z - z_0 \mid < \alpha - y_0$. We have

$$\delta_z^k f(z) = \sum_{m}^{\infty} (-1)^n (4n + 1)^k \exp\left[-(\alpha - ix_0)(4n + 1)^{\frac{1}{2}}\right] \frac{h_{2n}(z)}{A_{2n}}.$$

By Theorem 1.4

$$h_{2n}(x_0 + iy) = h_{2n}(iy) \exp\left[-ix_0(4n + 1)^{\frac{1}{2}}\right]\{1 + \eta_{2n}(x_0, y)(4n + 1)^{-\frac{1}{2}}\}.$$

Let us choose m so large that $\mid \eta_{2n}(x_0, y) \mid (4n + 1)^{-\frac{1}{2}} < \frac{1}{2}$ for $n \geqq m$ and $\frac{1}{2}\alpha \leqq y \leqq \alpha$. We have then

$$\begin{aligned} \mid \delta_z^k f(z_0) \mid &\geqq \Re[\delta_z^k f(z_0)] \\ &\geqq \tfrac{1}{2} \sum_{m}^{\infty} (-1)^n (4n + 1)^k \exp\left[-\alpha(4n + 1)^{\frac{1}{2}}\right] \frac{h_{2n}(iy_0)}{A_{2n}} \\ &> \tfrac{1}{4} \sum_{m}^{\infty} (4n + 1)^k \exp\left[-(\alpha - y_0)(4n + 1)^{\frac{1}{2}}\right], \end{aligned}$$

where the last inequality follows from Theorem 1.1. The sum of the last series exceeds its largest term. Now the function $u^k \exp[-au^{\frac{1}{2}}]$ reaches its maximum for $u = (2k/a)^2$, the maximum value being $(2k/(ea))^{2k}$, and the function exceeds one-half times its maximum value over an interval of length $O(k)$. It follows that

$$\mid \delta_z^k f(z_0) \mid > C\left(\frac{2k}{e}\right)^{2k} (\alpha - y_0)^{-2k},$$

where C is a positive constant independent of k. Hence

$$\liminf_{k\to\infty} k^{-2} \mid \delta_z^k f(z_0) \mid^{1/k} \geqq \left(\frac{2}{e}\right)^2 (\alpha - y_0)^{-2}.$$

For $k = 0$ we have in particular

$$\Re[f(x_0 + iy)] \geqq \tfrac{1}{4} \sum_m^\infty \exp\left[-(\alpha - y)(4n + 1)^{\frac{1}{2}}\right] \to \infty \text{ as } y \to \alpha.$$

It follows that $z = x_0 + i\alpha$ is a singular point of $f(z)$. Consequently $R = \alpha - y_0$ and for this particular function we have

$$\lim_{k\to\infty} k^{-2} \,|\, \delta_z^k f(z_0) \,|^{1/k} = \left(\frac{2}{eR}\right)^2. \tag{3.3.15}$$

In the case in which $z_0 = x_0 > 0$, we have to take into account the fact that $h_n(z)$ oscillates on the real axis. By formula (1.2.10)

$$|\, h_{2n}(x_0) - c_{2n}(x_0) \,| \leqq A_{2n}\left\{\exp \frac{x_0^3}{3(4n + 1)^{\frac{1}{2}}} - 1\right\}.$$

We choose an m first subject to the condition that the expression within braces shall be less than $\frac{1}{4}$ for $n \geqq m$. Further, using the same notation as above, we require that $|\, \eta_{2n}(x_0 , y) \,|\, (4n + 1)^{-\frac{1}{2}} < \frac{1}{2}$ for $n \geqq m$ and $\frac{1}{2}\alpha \leqq y \leqq \alpha$, where α is a given positive number. Let θ_n be the uniquely determined arc in the interval $[-\frac{1}{2}\pi, \frac{3}{2}\pi]$ to which $-(4n + 1)^{\frac{1}{2}}x_0$ is congruent modulo 2π, and define ϵ_n to be $(-1)^n$ if $-\frac{1}{4}\pi \leqq \theta_n \leqq \frac{1}{4}\pi$, and 0 otherwise.

If $\psi_n = \arg\{1 + \eta_{2n}(x_0 , y)(4n + 1)^{-\frac{1}{2}}\}$, then $-\frac{1}{6}\pi \leqq \psi_n \leqq \frac{1}{6}\pi$ for $n \geqq m$. Further, $\arg\{(-1)^n h_{2n}(x_0 + iy)\} = \theta_n + \psi_n$ lies between $\pm\frac{5}{12}\pi$ and $\cos(\theta_n + \psi_n) \geqq \sin \frac{1}{12}\pi > \frac{1}{4}$. After these preliminaries let us form

$$f(z) = \sum_m^\infty \epsilon_n \exp\left[-\alpha(4n + 1)^{\frac{1}{2}}\right] \frac{h_{2n}(z)}{A_{2n}}. \tag{3.3.16}$$

Here $R \geqq \alpha$ by virtue of the construction of the series. In order to prove that $R = \alpha$ it is enough to show that $z = x_0 + i\alpha$ is a singular point of $f(z)$. For $\frac{1}{2}\alpha \leqq y \leqq \alpha$ we have

$$\begin{aligned}\Re[f(x_0 + iy)] = &\sum_m^\infty |\, \epsilon_n \,| \exp\left[-\alpha(4n + 1)^{\frac{1}{2}}\right] |\, h_{2n}(iy) \,| \\ &\cdot |\, 1 + \eta_{2n}(x_0, y)(4n + 1)^{-\frac{1}{2}} \,| \cos(\theta_n + \psi_n) \frac{1}{A_{2n}} \\ &> \tfrac{1}{8} \sum_m^\infty |\, \epsilon_n \,| \exp\left[-\alpha(4n + 1)^{\frac{1}{2}}\right] \frac{|\, h_{2n}(iy) \,|}{A_{2n}} \\ &> \tfrac{1}{16} \sum_m^\infty |\, \epsilon_n \,| \exp\left[(y - \alpha)(4n + 1)^{\frac{1}{2}}\right] \to \infty \text{ as } y \to \alpha.\end{aligned}$$

This shows that $z = x_0 + i\alpha$ is a singular point and that $R = \alpha$. Further

$$\delta_x^k f(x_0) = \sum_m^\infty \epsilon_n (4n + 1)^k \exp\left[-\alpha(4n + 1)^{\frac{1}{2}}\right] \frac{h_{2n}(x_0)}{A_{2n}}$$

and all terms of this series are $\geqq 0$. In any interval of the form $(u_0 - ck, u_0 + ck)$, where $u_0 = (2k/\alpha)^2$ and c is a constant, the function $u^k \exp(-\alpha u^{\frac{1}{2}})$

exceeds $C(c)(2k/(e\alpha))^{2k}$. For a suitable choice of c, such an interval contains at least one integer $4n + 1$ such that $\epsilon_n \neq 0$ and for such an n we have

$$\epsilon_n \frac{h_{2n}(x_0)}{A_{2n}} \geqq \cos \theta_n - \tfrac{1}{4} > \tfrac{1}{2}\sqrt{2} - \tfrac{1}{4}.$$

Hence for large values of k

$$\delta_z^k f(x_0) > A \left(\frac{2k}{e\alpha}\right)^{2k},$$

where A is independent of k. Consequently

$$\liminf_{k\to\infty} k^{-2}[\delta_z^k f(x_0)]^{1/k} \geqq \left(\frac{2}{eR}\right)^2$$

and combining this result with formula (3.3.13), we get finally

$$\lim_{k\to\infty} k^{-2}[\delta_z^k f(x_0)]^{1/k} = \left(\frac{2}{eR}\right)^2. \tag{3.3.17}$$

We have consequently constructed counter examples valid for every preassigned set of finite values of z_0 and R. Thus we have proved

THEOREM 3.1. *If $f(z)$ is holomorphic in the circle $|z - z_0| < R$, then*

$$\limsup_{k\to\infty} k^{-2} |\delta_z^k f(z_0)|^{1/k} \leqq \left(\frac{2}{eR}\right)^2.$$

This estimate is the best possible of its kind in the sense that to every given set of finite values of z_0 and R there exist functions $f(z)$ for which lim sup *can be replaced by* lim *and the sign of equality holds.*

3.4. **Entire functions of δ_z.** After this study of the behavior of $\delta_z^k f(z)$ for large values of k, we are prepared to study entire functions of the operator δ_z. Let

$$G(w) = \sum_{k=0}^{\infty} g_k w^k \tag{3.4.1}$$

be a given entire function and let $f(z)$ be a given analytic function. We say that *the differential operator*

$$G(\delta_z) = \sum_{k=0}^{\infty} g_k \delta_z^k \tag{3.4.2}$$

is applicable to the function $f(z)$ if the series

$$G(\delta_z)f(z) \equiv \sum_{k=0}^{\infty} g_k \delta_z^k f(z) \tag{3.4.3}$$

converges at every point where $f(z)$ is holomorphic. We shall prove[7]

[7] Cf. H. Muggli [15], Satz I, A, p. 152. This paper contains references to previous literature.

THEOREM 3.2. *A necessary and sufficient condition in order that the differential operator $G(\delta_z)$ shall be applicable to every analytic function is that the entire function $G(w)$ is of order $\leqq \frac{1}{2}$ and that it is of the minimal type if the order equals $\frac{1}{2}$.*

Proof. We recall that the order of an entire function is defined by[8]

$$\rho = \limsup_{r\to\infty} \frac{\log \log M(r)}{\log r}, \quad \text{where} \quad M(r) = \max_{0\leq\theta<2\pi} |G(re^{i\theta})|,$$

and that $G(w)$ is of the minimal type of the order ρ if

$$\lim_{r\to\infty} r^{-\rho} \log M(r) = 0.$$

A necessary and sufficient condition in order that $G(w)$ shall be of the minimal type of order $\frac{1}{2}$ is that

$$\limsup_{k\to\infty} k^2 \,|\, g_k \,|^{1/k} = 0. \tag{3.4.4}$$

We can now prove that the condition of the theorem is sufficient. We know then that (3.4.4) holds. On the other hand, by Theorem 3.1

$$\limsup_{k\to\infty} k^{-2} \,|\, \delta_z^k f(z_0) \,|^{1/k} \leqq \left(\frac{2}{eR}\right)^2,$$

where R is the distance from z_0 to the nearest singular point of $f(z)$. Hence

$$\lim_{k\to\infty} |\, g_k \delta_z^k f(z_0) \,|^{1/k} = 0,$$

whence it follows that the series (3.4.3) is absolutely convergent at every point $z = z_0$ where $f(z)$ is holomorphic.[9]

Conversely, suppose that (3.4.3) converges for every analytic function $f(z)$, and suppose that (3.4.4) does not hold but instead

$$\limsup_{k\to\infty} k^2 \,|\, g_k \,|^{1/k} = (\tfrac{1}{2}eT)^2.$$

This is equivalent to assuming that $G(w)$ is of order $\frac{1}{2}$ and type T. If z_0 is a given point, we can find an analytic function $f(z)$ holomorphic in the circle $|\, z - z_0 \,| < R$ where $R < T$, and such that

$$\lim_{k\to\infty} k^{-2} \,|\, \delta_z^k f(z_0) \,|^{1/k} = \left(\frac{2}{eR}\right)^2.$$

Such functions were constructed in §3.3. If we substitute this particular function in the series in (3.4.3), the resulting series is necessarily divergent at $z = z_0$ because

$$\limsup_{k\to\infty} |\, g_k \delta_z^k f(z_0) \,|^{1/k} = \left(\frac{T}{R}\right)^2 > 1.$$

[8] For the properties of entire functions used here and in the following, see G. Valiron [25].

[9] Referring to footnote 6, we see that if $f(z)$ is holomorphic in a bounded domain D, having a positive distance from the set of singularities of $f(z)$, then the series in (3.4.3) converges uniformly in D. It follows that $G(\delta_z)f(z)$ is holomorphic in D. Cf. H. Muggli [15], Satz II, A for the operator d/dz.

Hence it is necessary that $T = 0$ in order that $G(\delta_z)$ shall be applicable to every analytic function. This completes the proof of the theorem.

3.5. **The case when $f(z)$ is an entire function.** In the two preceding sections we have assumed that $f(z)$ was an analytic function having finite singular points. It is clear that if $f(z)$ is an entire function we can let $R \to \infty$ in formula (3.3.13) obtaining

$$\lim_{k\to\infty} k^{-2} \, | \, \delta_z^k f(z_0) \, |^{1/k} = 0, \tag{3.5.1}$$

and, without any further information concerning $f(z)$, this is the best possible estimate.[10]

If we have information concerning the order and type of $f(z)$, better estimates can be found, however. We shall merely state the results here, detailed proofs and various extensions will be published elsewhere. To facilitate a brief formulation the following notation will be used. The letter E followed by one or two inequalities in parentheses will denote the class of all entire functions the orders and types of which satisfy the inequalities in question. Thus $E(\rho < \rho_1)$ is the class of entire functions of order $\rho < \rho_1$, whereas $E(\rho \leqq \rho_1, T \leqq T_1) = E(\rho < \rho_1) + E(\rho = \rho_1, T \leqq T_1)$ contains in addition all entire functions of order ρ_1 and type $\leqq T_1$. Instead of $E(\rho = \rho_1)$ and $E(\rho = \rho_1, T = T_1)$ we write simply $E(\rho_1)$ and $E(\rho_1, T_1)$, respectively.

With a given order ρ we associate *the conjugate order σ with respect to the operation δ_z* with the aid of the equations

$$\sigma = \tfrac{1}{2} \text{ if } \rho = \infty; \quad \frac{1}{\rho} + \frac{1}{2\sigma} = 1 \text{ if } \rho > 2; \quad \sigma = 1 \text{ if } 0 \leqq \rho \leqq 2. \tag{3.5.2}$$

We have then the following results.

THEOREM 3.3. (i) *If $f(z) \in E(\rho, T)$, $\rho > 2$, then*

$$\limsup_{k\to\infty} k^{-1/\sigma} \, | \, \delta_z^k f(z) \, |^{1/k} \leqq \left(\frac{2}{e}\right)^{1/\sigma} (\rho T)^{2/\rho}, \tag{3.5.3}$$

and for any given set of three values z, ρ, T there exists an $f(z)$ for which the sign of equality holds even for lim sup *replaced by* lim.

(ii) *If $f(z) \in E(2, T)$, then*

$$\limsup_{k\to\infty} k^{-1} \, | \, \delta^k f(z) \, |^{1/k} \leqq A + BT. \tag{3.5.4}$$

Limits can be obtained for the values of the constants A and B, but the best values are in doubt.

[10] From this estimate follows immediately that a sufficient condition that $G(\delta_z)$ shall be applicable to all entire functions $f(z)$ is that the entire function $G(w)$ is at most of the normal type of order $\frac{1}{2}$. It is not difficult to prove that this condition is also necessary. Cf. H. Muggli, Satz I, B.

(iii) *If $f(z) \epsilon E(\rho, T), 0 \leqq \rho < 2$, then*

$$\limsup_{k\to\infty} k^{-1} \mid \delta^k f(z) \mid^{1/k} \leqq C, \tag{3.5.5}$$

where C is an absolute constant, $1/e \leqq C \leqq 2^{\frac{1}{2}}$.

THEOREM 3.4.[11] *A necessary and sufficient condition that the differential operator $G(\delta_z)$ shall be applicable to every function $f(z)$ of the class $E(\rho \leqq \rho_1, T < \infty)$ is that the entire function $G(w)$ belongs to the class $E(\rho < \sigma_1) + E(\sigma_1, 0)$ provided $\rho_1 \geqq 2$. If $\rho_1 < 2$, the condition is merely sufficient, but it is necessary that $G(w) \epsilon E(\rho \leqq 1, T < 1)$.*

It seems likely that the differential operator $G(\delta_z)$ preserves the order of the entire function $f(z)$ when $f(z)$ and $G(w)$ belong to conjugate classes in the sense of Theorem 3.4, but this question requires further investigation.

Chapter 4. Regularity preserving factor sequence transformations of Hermitian series

4.1. **Problem and results.** Let $\mathfrak{A} = \{a_n\}$ be a given factor sequence such that

$$\lim_{n\to\infty} n^{-\frac{1}{2}} \log \mid a_n \mid = 0 \tag{4.1.1}$$

and consider the factor sequence transformation $F(z) = \mathfrak{A}[f(z)]$ which carries the Hermitian series

$$f(z) = \sum_{n=0}^{\infty} f_n h_n(z) \tag{4.1.2}$$

into

$$F(z) = \sum_{n=0}^{\infty} a_n f_n h_n(z). \tag{4.1.3}$$

Because of (4.1.1) the ordinate of convergence of the second series is not less than that of the first which we suppose to be positive.

What additional conditions should be imposed on $\mathfrak{A}$ in order that it shall always be possible to continue $\mathfrak{A}[f(z)]$ analytically along any finite path C from $z = 0$ along which $f(z)$ can be continued?

We shall give a couple of classes of such factor sequences in the following. We shall be guided largely by the analogy with the theory of power series, Dirichlet series, and Laplace-Stieltjes integrals. A classical theorem of L. Leau [14] which has been given its final and most general formulation by A. Ostrowski [17] states that *if $g(w)$ is defined at $w = n$ ($n = 0, 1, 2, \cdots$) and is holomorphic in some neighborhood of the point at infinity, then the function defined by the element $\sum_{n=0}^{\infty} g(n)w^n$ can be continued analytically along any path C on the Riemann surface*

[11] Cf. H. Muggli [15], Satz I, C and D for the operator d/dz. There is good analogy only in the first case.

of $\log(1-w)$ *which, starting from* $w = 0$, *does not pass through any of the points* $0, 1,$ *or* ∞. On the basis of this theorem Ostrowski[12] proved that *if*

$$\varphi(w) = \sum_{n=0}^{\infty} c_n w^n, \qquad \Phi(w) = \sum_{n=0}^{\infty} c_n g(n) w^n,$$

then $\Phi(w)$ *can be continued analytically along any path, starting from* $w = 0$, *which does not pass through either* $w = 0$ *or* $w = \infty$ *along which* $\varphi(w)$ *can be continued.*

We shall prove the following analogue of this theorem in §4.2.

THEOREM 4.1. *Let* $g(w)$ *be defined for* $w = n$ $(n = 0, 1, 2, \cdots)$ *and holomorphic in some neighborhood of the point at infinity. Let*

$$f(z) = \sum_{n=0}^{\infty} f_n h_n(z), \qquad F(z) = \sum_{n=0}^{\infty} f_n g(n) h_n(z). \tag{4.1.4}$$

Then the two series have the same ordinate of convergence τ, *supposed to be positive, and* $F(z)$ *can be continued analytically along any finite path along which* $f(z)$ *can be continued.*

In other words, $\{g(n)\}$ *is a regularity preserving factor sequence for Hermitian series if* $g(w)$ *is holomorphic at* $w = \infty$. Another example of such sequences for power series is connected with the classical theorem of S. Wigert [33] according to which *the series* $\sum_0^{\infty} G(n)w^n$ *defines an entire function of* $(1-w)^{-1}$, *if* $G(w)$ *is an entire function of order one and minimal type.* The combination of this result with Hadamard's multiplication theorem gives a result, the prototype of which is due to G. Faber [6] and which can nowadays be formulated as follows.[13] *If* $G(w)$ *is an entire function of order one and minimal type and if*

$$\varphi(w) = \sum_{n=0}^{\infty} c_n w^n, \qquad \Phi(w) = \sum_{n=0}^{\infty} c_n G(n) w^n,$$

then $\Phi(w)$ *can be continued analytically along any finite path along which* $\varphi(w)$ *can be continued.* The analogue of this theorem for Hermitian series reads as follows.

THEOREM 4.2. *Let* $G(w)$ *be an entire function of order* $\leqq \frac{1}{2}$ *and of minimal type if the order equals* $\frac{1}{2}$. *Let*

$$f(z) = \sum_{n=0}^{\infty} f_n h_n(z), \qquad F(z) = \sum_{n=0}^{\infty} f_n G(n) h_n(z), \tag{4.1.5}$$

where the ordinate of convergence of the first series shall be $\tau > 0$. *Then the ordinate of convergence of the second series is* $\geqq \tau$, *and* $F(z)$ *can be continued analytically along any path in the finite part of the plane along which* $f(z)$ *can be continued.*[14]

[12] Loc. cit., pp. 719–720.

[13] This theorem is a special case of Cramér's theorem, quoted below, obtained by putting $T = 0$ and $w = e^{-s}$.

[14] Theorem 4.2 is a special case of Theorem 4.3.

This theorem will be proved in §4.3. It expresses that $\{G(n)\}$ *is a regularity preserving factor sequence for Hermitian series if* $G(w)$ *is at most of the minimal type of order* $\frac{1}{2}$.

Faber's original theorem has been generalized many times. One of the most interesting generalizations is the theorem of H. Cramér [3, 4] which can be formulated as follows. *Let* $G(w)$ *be an entire function of order one and type* T. *Let*

$$d(s) = \sum_{n=1}^{\infty} c_n e^{-\lambda_n s}, \qquad D(s) = \sum_{n=1}^{\infty} c_n G(\lambda_n) e^{-\lambda_n s}$$

be two Dirichlet series with $0 < \lambda_n < \lambda_{n+1}$, $\lambda_n \to +\infty$, *of which the first shall have a finite abscissa of convergence* σ_0. *Then the second series converges for* $\Re(s) > \sigma_0 + T$, *and* $D(s)$ *can be continued analytically along any part* C *in the finite part of the plane, starting from* $s = \sigma_0 + T + 1$ *say, such that* $d(s)$ *is holomorphic in a circle of radius* $> T$ *about every point of* C. In view of the close analogy between Hermitian series and Dirichlet series with exponents $\pm(2n+1)^{\frac{1}{2}}$ it is not surprising that Cramér's theorem has an analogue for Hermitian series, viz.,

THEOREM 4.3. *Let* $G(w)$ *be an entire function of order* $\frac{1}{2}$ *and type* T. *Let*

$$f(z) = \sum_{n=0}^{\infty} f_n h_n(z), \qquad F(z) = \sum_{n=0}^{\infty} f_n G(2n+1) h_n(z) \tag{4.1.6}$$

and suppose that the ordinate of convergence of the first series is $\tau > T$. *Then the ordinate of convergence of the second series is* $\geqq \tau - T$, *and* $F(z)$ *can be continued analytically along any path* C *in the finite part of the plane, starting from the point* $z = 0$ *say, such that* $f(z)$ *is holomorphic in a circle of radius* $> T$ *about every point of* C.

This theorem will be proved in §4.4.

Ostrowski has given several extensions of his Leau-Hadamard theorem quoted above. The most general statement actually proved in his paper is the following theorem.[15]

THEOREM LHO. *Let* $A(u)$ *and* $B(u)$ *be functions of bounded variation in every finite interval* $[a, \omega]$ *and* $[b, \omega]$, *respectively, where* a *and* b *are fixed positive quantities. Let*

$$f(s) = \int_a^{\infty} e^{-su}\, dA(u) \tag{4.1.7}$$

be convergent for $\Re(s) > \sigma_0 > -\infty$. *Let* $g(w)$ *be continuous for real values of* w, $a \leqq w < \infty$, *and let* $g(w)$ *be defined by the convergent integral*

$$g(w) = \int_b^{\infty} w^{-t}\, dB(t) \tag{4.1.8}$$

[15] Op. cit., p. 738.

for $|w| > \beta$. *Then the integral*

$$\varphi(s) = \int_a^\infty g(u)e^{-su}\,dA(u) \tag{4.1.9}$$

converges for $\Re(s) > \sigma_0$ *and can be continued analytically out of the half plane of convergence along any path in the finite part of the plane, which does not intersect itself, along which* $f(s)$ *can be continued. If, in addition,* $g(w)$ *is single-valued and analytic in some neighborhood of* $w = \infty$, *the condition on non-self-intersection can be omitted.*

The main point in the proof consists in establishing the representation

$$\varphi(s) = \int_0^\infty \frac{G(v)}{v} f(s + v)\,dv \tag{4.1.10}$$

with

$$G(w) = \int_b^\infty \frac{w^t}{\Gamma(t)}\,dB(t). \tag{4.1.11}$$

Even this theorem admits of a generalization to the Hermitian case. But such a generalization requires as a first step the development of a fairly extensive theory of *Hermite-Stieltjes integrals* of the form

$$f(z) = \int_0^\infty h_u(z)\,dA(u), \tag{4.1.12}$$

where $h_u(z)$ is some suitable interpolation function of the sequence $\{h_n(z)\}$. The choice of $h_u(z)$ can be made in infinitely many different ways, but the most natural choice is perhaps given by

$$h_u(z) = -\frac{\Gamma(u+1)}{2\pi i} e^{-\frac{1}{2}z^2} \int_C e^{-2zt-t^2}(-t)^{-u-1}\,dt, \tag{4.1.13}$$

where $-\pi < \arg(-t) < \pi$ and the integral surrounds the positive real axis in the positive sense.[16] In other words, we define $h_u(z)$ as $2^{\frac{1}{2}u}D_u(2^{\frac{1}{2}}z)$, where $D_u(w)$ is the function of Whittaker. The theory of the resulting Hermite-Stieltjes integrals is largely analogous to that of Laplace-Stieltjes integrals. If in (4.1.9) we replace e^{-su} by $h_u(z)$, a formula of type (4.1.10) still results where, however, $f(s + v)$ has to be replaced by an Abel transform of $f(z)$, viz.,

$$f(z, v) = \int_a^\infty e^{-vu}h_u(z)\,dA(u), \tag{4.1.14}$$

and the main difficulty in extending the LHO theorem to Hermitian series lies in the discussion of the behavior of $f(z, v)$ as $v \to 0$. The details of this investigation will have to be postponed to some other occasion.

[16] Cf. E. T. Whittaker [32], p. 422.

It is sometimes useful to keep in mind that *the regularity preserving factor sequences of Hermitian series form a ring* inasmuch as $\{a_n + b_n\}$ and $\{a_n b_n\}$ are clearly regularity preserving if $\{a_n\}$ and $\{b_n\}$ have this property.

4.2. **Proof of Theorem** 4.1. By assumption $g(w)$ is holomorphic at infinity. Let us expand $g(w)$ in a Laurent series about the point $w = -\frac{1}{2}$ and suppose that

$$g(w) = \sum_{k=0}^{\infty} g_k(2w + 1)^{-k}, \tag{4.2.1}$$

the series being convergent for $|w + \frac{1}{2}| > R$. Let

$$f(z) = \sum_{n=0}^{\infty} f_n h_n(z)$$

have $\tau > 0$ as ordinate of convergence. We can assume without restricting the generality that $f_0 = f_1 = \cdots = f_m = 0, f_{m+1} \neq 0$, where $m + 1 > R$. If this is not the case originally, we have merely to omit a finite number of terms at the beginning of the series. The resulting new series has the same ordinate of convergence and the same finite singularities, if any, as the old series.

We can choose an R_1, $R < R_1 < m + 1$ and a finite M such that

$$|g_k| < M(2R_1)^k \tag{4.2.2}$$

for all k. The double series

$$\sum_{n=0}^{\infty} \sum_{k=0}^{\infty} f_n g_k (2n + 1)^{-k} h_n(z)$$

is easily shown to be absolutely convergent for any fixed z in the strip $-\tau < y < \tau$. If we sum first for k and then for n, we get

$$\sum_{n=0}^{\infty} f_n g(n) h_n(z) \equiv F(z).$$

Changing the order of summation, we obtain

$$F(z) = \sum_{k=0}^{\infty} g_k \delta_z^{-k} f(z), \tag{4.2.3}$$

$$\delta_z^{-k} f(z) = \sum_{n=0}^{\infty} f_n (2n + 1)^{-k} h_n(z). \tag{4.2.4}$$

These series can be summed with the aid of the formulas of §3.2. Taking $z_0 = 0$ in (3.2.9), we get

$$\begin{aligned} \delta_z^{-k} f(z) = W_{k,1}(z) &+ \int_0^z G_1(z, t) W_{k,2}(t)\, dt + \cdots \\ &+ \int_0^z G_{k-1}(z, t) W_{k,k}(t)\, dt + \int_0^z G_k(z, t) f(t)\, dt. \end{aligned} \tag{4.2.5}$$

Here $W_{k,m}(z)$, $m \leqq k$, is that solution of

$$w'' - z^2 w = 0 \tag{4.2.6}$$

which is determined by the initial values

$$W_{k,m}(0) = [\delta_z^{m-1}\delta_z^{-k} f(z)]_{z=0},$$

$$W'_{k,m}(0) = \left[\frac{d}{dz}\,\delta_z^{m-1}\delta_z^{-k} f(z)\right]_{z=0}.$$

But

$$\delta_z^{m-1}\delta_z^{-k} f(z) = \delta_z^{-(k-m+1)} f(z)$$

for the values of k and m under consideration. Let us therefore introduce a sequence of solutions of (4.2.6), $\{W_j(z)\}$ ($j = 1, 2, 3, \cdots$), determined by the initial conditions

$$\begin{aligned} W_j(0) &= \sum_{n=0}^{\infty} f_n(2n+1)^{-j} h_n(0) \equiv F_j, \\ W'_j(0) &= \sum_{n=0}^{\infty} f_n(2n+1)^{-j} h'_n(0) \equiv F'_j. \end{aligned} \tag{4.2.7}$$

Then

$$\begin{aligned} \delta_z^{-k} f(z) = W_k(z) &+ \int_0^z G_1(z, t) W_{k-1}(t)\, dt + \cdots \\ &+ \int_0^z G_{k-1}(z, t) W_1(t)\, dt + \int_0^z G_k(z, t) f(t)\, dt. \end{aligned} \tag{4.2.8}$$

Let us multiply both sides of this equation by g_k and sum for k. We shall show that the resulting double series is absolutely convergent at least in the strip $-\tau < y < \tau$. We imagine the series written as a triangular array with $g_0 f(z)$ at the apex, the two terms of $\delta^{-1} f(z)$ in the second row, the three terms of $\delta^{-2} f(z)$ in the third, etc. Then we collect the terms involving $f(z)$ or $f(t)$ which form the right border of the array, and sum the remaining terms by columns. The result can be written

$$F(z) = g_0 f(z) + \int_0^z \Gamma(z, t) f(t)\, dt + \mathfrak{W}_0(z) + \sum_{n=1}^{\infty} \int_0^z G_n(z, t)\mathfrak{W}_n(t)\, dt. \tag{4.2.9}$$

Here

$$\Gamma(z, t) = \sum_{k=1}^{\infty} g_k G_k(z, t), \tag{4.2.10}$$

$$\mathfrak{W}_n(z) = \sum_{k=1}^{\infty} g_{k+n} W_k(z). \tag{4.2.11}$$

We shall prove the absolute convergence of these series and start with the last one. From the definition of $W_k(z)$ follows

$$W_k(z) = 2^{-\frac{1}{2}}\Gamma(\tfrac{3}{4})F_k w_1(z) + 2^{-\frac{1}{2}}\Gamma(\tfrac{1}{4})F_k' w_2(z), \tag{4.2.12}$$

where $w_1(z)$ and $w_2(z)$ are the solutions of (4.2.6) defined by formula (3.2.4). Now (4.2.7) shows that

$$| F_k |, | F_k' | = O\{(2m + 3)^{-k}\}. \tag{4.2.13}$$

Combining (4.2.2) with the last two formulas and using trivial estimates of the Bessel functions, we get

$$| \mathfrak{W}_n(z) | \leqq C(2R_1)^n \exp (\tfrac{1}{2} | z |^2) \sum_{k=0}^{\infty} \left(\frac{2R_1}{2m + 3}\right)^k$$

or

$$| \mathfrak{W}_n(z) | \leqq K(2m + 2)^{n+1} \exp (\tfrac{1}{2} | z |^2), \tag{4.2.14}$$

where K is an absolute constant.

From the definition of $G_1(z, t)$ follows that we can find a constant A such that

$$| G_1(z, t) | \leqq Ae^{p^2}, \qquad p = \max (| z |, | t |).$$

Using standard estimates we then find that

$$| G_n(z, t) | \leqq A^n e^{np^2} \frac{| z - t |^{n-1}}{(n - 1)!}. \tag{4.2.15}$$

It follows that

$$| \Gamma(z, t) | \leqq M(2m + 2)Ae^{p^2} \exp \{(2m + 2)Ae^{p^2} | z - t |\}. \tag{4.2.16}$$

Formulas (4.2.14) and (4.2.15) suffice to prove that

$$\mathfrak{W}(z) \equiv \mathfrak{W}_0(z) + \sum_{n=1}^{\infty} \int_0^z G_n(z, t)\, \mathfrak{W}_n(t)\, dt \tag{4.2.17}$$

is an entire function. They also show that the terms under the main diagonal in the triangular array form an absolutely convergent series in the finite part of the plane.

Let us now consider

$$\mathfrak{F}(z) \equiv g_0 f(z) + \int_0^z \Gamma(z, t) f(t)\, dt. \tag{4.2.18}$$

Formula (4.2.16) shows that $\Gamma(z, t)$ is an entire function in z as well as in t. It follows that $\mathfrak{F}(z)$, which to start with is defined merely in the strip $-\tau < y < \tau$, can be continued analytically outside of this strip along any finite rectifiable path along which $f(z)$ can be continued. Since finally

$$F(z) = \mathfrak{F}(z) + \mathfrak{W}(z), \tag{4.2.19}$$

the proof of Theorem 4.1 is complete.

4.3. **Proof of Theorem** 4.2. This theorem is a fairly simple consequence of Theorem 3.2. Let

$$f(z) = \sum_{n=0}^{\infty} f_n h_n(z)$$

have $\tau > 0$ as its ordinate of convergence. The series

$$F(z) = \sum_{n=0}^{\infty} f_n G(n) h_n(z) \tag{4.3.1}$$

evidently has a sense. The assumptions on $G(w)$ imply that

$$\limsup_{n\to\infty} n^{-1} \log |G(n)| \leqq 0,$$

so that the ordinate of convergence of (4.3.1) is $\geqq \tau$. Let

$$G(\tfrac{1}{2}(w-1)) = \sum_0^{\infty} g_n w^n. \tag{4.3.2}$$

Since this function is also an entire function of w of order $\leqq \frac{1}{2}$ and of minimal type if its order equals $\frac{1}{2}$, Theorem 3.2 shows that the operator $G(\frac{1}{2}(\delta_z - 1))$ is applicable to $h_n(z)$ for every n and

$$\begin{aligned} G(\tfrac{1}{2}(\delta_z - 1))\cdot h_n(z) &= \sum_0^{\infty} g_k \delta_z^k h_n(z) \\ &= \sum_0^{\infty} g_k (2n+1)^k h_n(z) = G(n) h_n(z). \end{aligned}$$

Consequently

$$G(\tfrac{1}{2}(\delta_z - 1))\cdot\left\{\sum_0^N f_n h_n(z)\right\} = \sum_0^N f_n G(n) h_n(z)$$

for every N. As $N \to \infty$, the series on the left converges uniformly to $f(z)$ in the domain $-1/\epsilon \leqq x \leqq 1/\epsilon$, $-\tau + \epsilon \leqq y \leqq \tau - \epsilon$, and the series on the right converges uniformly to $F(z)$ in the same domain. Hence

$$G(\tfrac{1}{2}(\delta_z - 1))\cdot f(z) = F(z) \tag{4.3.3}$$

within the strip of convergence of the two series. But Theorem 3.2 shows that the left side exists as a holomorphic function of z in any circle within which $f(z)$ is holomorphic. Consequently $F(z)$ can be continued analytically along any finite path along which $f(z)$ can be continued. This completes the proof of Theorem 4.2.

4.4. **Proof of Theorem** 4.3. We proceed as in the preceding section. By assumption $G(w)$ is an entire function of order $\frac{1}{2}$ and type T. Hence

$$\limsup_{n\to\infty} (2n+1)^{-1} \log |G(2n+1)| \leqq T. \tag{4.4.1}$$

This implies that if τ, by assumption greater than T, is the ordinate of convergence of the series

$$f(z) = \sum_0^\infty f_n h_n(z),$$

then the series

$$F(z) = \sum_0^\infty f_n G(2n + 1)h_n(z) \tag{4.4.2}$$

has an ordinate of convergence $\geqq \tau - T$.

We have next to prove that

$$G(\delta_z)f(z) = F(z) \tag{4.4.3}$$

for $-(\tau - T) < y < \tau - T$. The proof follows the same lines as that of the preceding section and can be left to the reader.

Now suppose that $f(z)$ can be continued analytically along a finite path C leading from $z = 0$ to $z = z_0$ and that if ζ is any point of C, then $f(z)$ is holomorphic in the circle $|z - \zeta| < R(\zeta)$, where $R(\zeta) > T$ for every ζ on C. If

$$G(w) = \sum_0^\infty \gamma_n w^n,$$

then

$$\{G(\delta_z)f(z)\}_{z=\zeta} = \sum_0^\infty \gamma_n \delta_\zeta^n f(\zeta) \tag{4.4.4}$$

is absolutely and uniformly convergent on C. Indeed, by formula (3.3.13)

$$\limsup_{n\to\infty} n^{-2} | \delta_\zeta^n f(\zeta) |^{1/n} \leqq \left(\frac{2}{e}\right)^2 [R(\zeta)]^{-2},$$

whereas

$$\limsup_{n\to\infty} n^2 | \gamma_n |^{1/n} = (\tfrac{1}{2}e)^2 T^2,$$

since $G(w)$ is of order $\frac{1}{2}$ and type T. These two estimates imply the absolute and uniform convergence of (4.4.4) since $R(\zeta) \geqq T + \epsilon$ for some fixed positive ϵ on the closed bounded set C. Hence $F(z)$ is holomorphic on C.

We conclude that $F(z)$ can be continued analytically along any path of the character stated in Theorem 4.3 and this theorem is proved.

Theorem 4.3 is the best of its kind in a certain sense. Thus if $\tau = T$ it may happen that the transformed series does not represent an analytic function at all. In order to construct such an example, let us consider an arbitrary function $\varphi(x) \in L_2(-\infty, \infty)$ and let

$$\varphi(x) \sim \sum_0^\infty \varphi_n h_n(x),$$

where, as is well known, the only condition the coefficients have to satisfy is the convergence of the series

$$\sum_0^\infty 2^n n! \, |\varphi_n|^2.$$

Now define

$$f(z, T) = \sum_0^\infty \varphi_n \exp[-T(2n+1)^{\frac{1}{2}}] \, h_n(z).$$

This is easily seen to represent a holomorphic function in the strip $-T < y < T$. Then take

$$G(w) = \cosh [Tw^{\frac{1}{2}}].$$

This is an entire function of order $\frac{1}{2}$ and type T and the corresponding transformed series

$$\sum_0^\infty \varphi_n \cosh [T(2n+1)^{\frac{1}{2}}] \exp[-T(2n+1)^{\frac{1}{2}}] \, h_n(x)$$

represents

$$\tfrac{1}{2}\varphi(x) + \tfrac{1}{2}f(x, 2T),$$

i.e., is not an analytic function.

If $\tau > T$ it may very well happen that the lines $y = \pm(\tau - T)$ are natural boundaries of $F(z)$. This will be the case in particular if $f(z)$ is a suitably chosen gap series, having the strip $-\tau < y < \tau$ as its domain of existence and $G(w) = \cosh [Tw^{\frac{1}{2}}]$.

Chapter 5. Singularities of Hermitian series

5.1. **Objectives.** In the present chapter we shall study the singularities of functions defined by Hermitian series, in particular, the singularities on the lines of convergence. It is natural that in this first investigation of the field we can at most hope to scratch the surface. But we shall prove some of the obviously basic theorems including some gap theorems. We shall be guided mainly by the close analogy between Hermitian series on one hand and power and Dirichlet series on the other. Of the vast literature on singularities of Dirichlet series, we make explicit mention of two fundamental papers by F. Carlson and E. Landau [2] and by O. Szász [23] which have been found particularly suggestive in connection with this investigation.

The location and nature of the singularities of a function defined by a power series are determined by the asymptotic properties of the sequence of coefficients. The efforts of finding relations between coefficients and singularities have been directed along several different lines. Thus, for instance, the known properties of a given analytic function $g(w)$ have been utilized in many important cases to determine the singularities of the power series $\sum g(n)w^n$. Some such instances were mentioned in §4.1 above. While this line of attack will undoubtedly turn

out to be fruitful also in the case of Hermitian series, no attempt will be made in the present paper to utilize this rather special method.

More general modes of attack are based either on what might be called the degree of dispersion of the coefficients or on the presence of gaps in the sequence. In both cases it is a question of frequency considerations, in the first case referring to the arguments, in the second case to the absolute values of the coefficients. Thus if all but a finite number of coefficients of a power series have the same argument, the point $z = R$, where R is the radius of convergence of the series, is singular. The same result is true if all but a finite number of the arguments are restricted to a sector of opening less than π, and this generalized theorem extends also to Dirichlet series and Laplace-Stieltjes integrals. We shall prove several analogues of this theorem for Hermitian series. Reality considerations on the imaginary axis show that z^n in the power series case does not correspond to $h_n(z)$ in the Hermitian case but to $i^{-n}h_n(z)$. Thus *if* $i^n f_n \geqq 0$ *for all large* n, *then* $z = \tau i$ *is a singular point of the function defined by the series.* For series with real terms the situation is not quite so simple. If the terms of even order (or those of odd order) have alternating signs, and if the suppression of the remaining terms does not alter the ordinate of convergence, then at least one of the points $\pm\, \tau i$ must be singular, but we can not tell for sure which point is singular in general.

Thus lack of dispersion in the arguments of the coefficients of a Hermitian series places at least one singular point on the imaginary axis. In view of the corresponding situation for Dirichlet series, it is not surprising that this property persists if some dispersion is allowed provided the frequency is sufficiently low.

On the other hand, it is well known from the theory of power series that strong dispersion of the arguments of the coefficients restricts the domain of existence of the function to the circle of convergence of the series. We illustrate the corresponding phenomenon for Hermitian series by showing that suitable changes of the signs of the coefficients of a given Hermitian series lead to non-continuable series.

The classical method of constructing non-continuable power series and Dirichlet series is by means of gaps. This method also works for Hermitian series. There is an obvious analogy between Hermitian series on one hand and Dirichlet series with exponents $\pm\, (2n + 1)^{\frac{1}{2}}$ on the other, the deeper reason of which we shall endeavor to bring out in Chapter 6. While this relationship is so close that one can safely predict much of the behavior of Hermitian series from the knowledge of the behavior of the associated Dirichlet series and use essentially the same methods to study both, the author is not able at present to derive results on singularities of one series from known theorems on singularities of the other. Thus we get our gap theorems not from classical gap theorems but from classical methods used to prove gap theorems. According to a famous dictum of E. H. Moore there should exist an abstract theory covering both these instances. The character and structure of this abstract theory is beyond my knowledge.

5.2. **Some auxiliary theorems on entire functions.** In the following sections we shall need some results concerning entire functions of order $\frac{1}{2}$ and minimal type. These theorems are largely analogous to classical theorems on entire functions of order one and minimal type which are already in the literature and could, at least in part, be derived from such theorems by simple transformations. Other needed results are not directly obtainable from the literature and in order to facilitate the reading of this paper the author has included complete proofs of all the required auxiliary theorems.

In the following $\{a_n\}$ is a given monotone increasing sequence of real positive quantities and $N(t)$ *is the number of a's less than or equal to t.* We introduce the two functions

$$\omega(t) = \frac{t^{\frac{1}{2}}}{N(t)}, \qquad \omega^*(t) = \inf_{u \geq t} \omega(u). \tag{5.2.1}$$

The basic assumption throughout this section is

HYPOTHESIS A. $\omega(t) \to \infty$ *with* t.

The corresponding entire function

$$G(w) = \prod_{n=1}^{\infty} \left\{1 - \frac{w}{a_n}\right\} \tag{5.2.2}$$

is then of order $\rho \leq \frac{1}{2}$ and of minimal type if $\rho = \frac{1}{2}$. Further $| G(re^{i\theta}) | \leq G(-r)$ for all θ. Our first problem is to find lower estimates of

$$\log | G(u) | = \int_0^\infty \log \left| 1 - \frac{u}{t} \right| dN(t) \tag{5.2.3}$$

for large positive values of u outside of fixed intervals around the zeros. We prove first

LEMMA 5.1. *Let* $\epsilon(u) > 0$ *and* $\epsilon(u) \to 0$ *as* $u \to \infty$ *but in such a manner that* $\epsilon(u)\omega(u) \to \infty$, *while* $\epsilon(u)\omega(u)[\omega(2u)]^{-1} \to 0$,[17] *and put*

$$\vartheta(u) = u \exp [-\epsilon(u)\omega(u)]. \tag{5.2.4}$$

Then

$$\lim_{u \to \infty} u^{-\frac{1}{2}} \left\{ \int_0^{u - \vartheta(u)} + \int_{u + \vartheta(u)}^\infty \right\} \log \left| 1 - \frac{u}{t} \right| dN(t) = 0. \tag{5.2.5}$$

Proof. Since the superior limit of the quantity under consideration in (5.2.5) is 0 by virtue of Hypothesis A, it is enough to prove that the inferior limit is ≥ 0.[18] To simplify we write $\vartheta(u) = \vartheta$. Then

[17] The choice $\epsilon(u) = [\omega(2u)]^{\frac{1}{2}}[\omega(u)]^{-1}$ satisfies these conditions. The last condition on $\epsilon(u)$ enters in the discussion of the integral from $u + \vartheta$ to $2u$ which is omitted below.

[18] Note that the multiplier of $u^{-\frac{1}{2}}$ in (5.2.5) is dominated by $\log | G(u) |$ which in turn is dominated by the integral from $a_1 - 0$ to $\frac{1}{2}u$ of the integrand in (5.2.3). A simple integration by parts shows that this integral is less than twice the integral from a_1 to $\frac{1}{2}u$ of $N(t)/t$ which is $o(u^{\frac{1}{2}})$ by Hypothesis A. This proves the statement concerning the superior limit.

$$u^{-\frac{1}{2}}\int_{0}^{u-\vartheta}\cdots > u^{-\frac{1}{2}}\int_{\frac{1}{2}u}^{u-\vartheta}\cdots > u^{-\frac{1}{2}}\log\left|1-\frac{u}{u-\vartheta}\right|\int_{\frac{1}{2}u}^{u-\vartheta} dN(t)$$

$$> u^{-\frac{1}{2}}\log\frac{\vartheta}{u-\vartheta}N(u) > u^{-\frac{1}{2}}\log\frac{\vartheta}{u}N(u)$$

$$= u^{-\frac{1}{2}}[-\epsilon(u)\omega(u)]N(u) = -\epsilon(u) \to 0$$

as $u \to \infty$, and a similar argument applies to the integral from $u + \vartheta$ to $2u$. Further

$$u^{-\frac{1}{2}}\int_{2u}^{\infty}\cdots > -2u^{\frac{1}{2}}\int_{2u}^{\infty}\frac{dN(t)}{t} > -2u^{\frac{1}{2}}\int_{2u}^{\infty}\frac{N(t)}{t^2}\,dt$$

$$> -2\frac{u^{\frac{1}{2}}}{\omega^*(u)}\int_{2u}^{\infty} t^{-\frac{3}{2}}\,dt = -\frac{2\sqrt{2}}{\omega^*(u)} \to 0.$$

This completes the proof of the lemma.

It is consequently enough to investigate the behavior of

$$\gamma(u) = \int_{u-\vartheta(u)}^{u+\vartheta(u)} \log\left|1-\frac{u}{t}\right| dN(t) \tag{5.2.6}$$

in the following. We shall determine additional conditions on $N(t)$ in order that $u^{-\frac{1}{2}}\gamma(u) \to 0$ as $u \to \infty$ in a specified manner. The following hypotheses are sufficient for this purpose.[19]

Hypothesis B. *$\omega(t)/\log t \to \infty$ with t.*

Hypothesis C. *Hypothesis* A *holds and, in addition, there exists a constant $c > 0$ such that*

$$\int_{u}^{u+v} dN(t) < 2 \quad \textit{for} \quad 0 < v \leq cu^{\frac{1}{2}} \quad \textit{and all} \quad u \geq 1. \tag{5.2.7}$$

Condition (5.2.7) implies and is implied by

$$a_{n+1} - a_n \geq ca_n^{\frac{1}{2}}. \tag{5.2.8}$$

Lemma 5.2.[20] *Let $h > 0$ and let I_h denote the open set on the positive real axis which remains after the intervals $[a_n - h, a_n + h]$ $(n = 1, 2, 3, \cdots)$ have been deleted. If $\{a_n\}$ satisfies either of Hypotheses* B *or* C, *and if $u \to \infty$ through any set of values in I_h, then* $\lim u^{-\frac{1}{2}} \log |G(u)| = 0$.

Proof. By virtue of Lemma 5.1 it is enough to prove that $\liminf u^{-\frac{1}{2}}\gamma(u) \geq 0$ when $u \to \infty$ in I_h, the superior limit being known to be 0. Suppose now that

[19] Hypothesis B does not seem to have any analogue in the existing literature on Dirichlet series whereas Hypothesis C is the analogue of Fabry's gap condition for power series [8, 9] and of the conditions of Carlson and Landau [2] and Szász [23] for Dirichlet series.

[20] For the case of the analogue of Hypothesis C, cf. O. Szász [23], Hilfsatz 3, pp. 102–104.

$u \in I_h$. This means that the distance from u to the nearest point of discontinuity of $N(t)$ exceeds h, so that

$$\int_{u-\vartheta}^{u+\vartheta} = \int_{u-\vartheta}^{u-h} + \int_{u+h}^{u+\vartheta} > -\log \frac{u}{h} \cdot N(u) - \log \frac{2u}{h} \cdot N(2u).$$

If Hypothesis B holds, the limit of the last member divided by $u^{\frac{1}{2}}$ is zero, and the lemma is proved.

If Hypothesis C holds instead, a more elaborate discussion is necessary.[21] Suppose that the points a_n situated in the interval $[u - \vartheta, u + \vartheta]$ are $a_k < a_{k+1} < \cdots < a_l < u < a_{l+1} < \cdots < a_m$. Using (5.2.8) we find for $n < l$

$$\begin{aligned} u - a_n &= (u - a_l) + (a_l - a_{l-1}) + \cdots + (a_{n+1} - a_n) \\ &> h + ca_{l-1}^{\frac{1}{2}} + \cdots + ca_n^{\frac{1}{2}} > c(l-n)a_n^{\frac{1}{2}}, \end{aligned}$$

whereas for $n > l + 1$

$$a_n - u > h + ca_{l+1}^{\frac{1}{2}} + \cdots + ca_n^{\frac{1}{2}} > c(n - l - 1)u^{\frac{1}{2}}.$$

Hence

$$\begin{aligned} \gamma(u) &= \int_{u-\vartheta}^{u+\vartheta} \log \left| 1 - \frac{u}{t} \right| dN(t) \\ &> 2 \log h + (m - k - 1) \log c + \log [(l-k)!\,(m-l-1)!] \\ &\quad + \frac{1}{2} \sum_k^{l-1} \log a_n + \frac{1}{2}(m - l - 1) \log u - \sum_k^m \log a_n. \end{aligned}$$

Here

$$(l-k)!(m-l-1)! > 2^{k+1-m}(m-k-1)!$$

and

$$\begin{aligned} \frac{1}{2} \sum_k^{l-1} \log a_n + \frac{1}{2}(m - l - 1) \log u - \sum_k^m \log a_n \\ > \tfrac{1}{2}(m - 2l + k - 1) \log u - (m + 1 - l) \log (u + \vartheta) \\ > -\tfrac{1}{2}(m - k + \tfrac{3}{2}) \left[\log u + 2 \log \left(1 + \frac{\vartheta}{u} \right) \right]. \end{aligned}$$

Thus

$$\gamma(u) > (m - k + 1) \left\{ \log (m - k + 1) + \log \frac{c}{2e} - \frac{1}{2} \log u - \log \left(1 + \frac{\vartheta}{u} \right) \right\} + O(\log u).$$

[21] This argument goes back to G. Faber [7], pp. 72–73. See also A. Pringsheim [21], pp. 87–91, who simplified the discussion.

Here the right side is certainly negative since $\gamma(u) < 0$. Noting that $m - k + 1 = N(u + \vartheta) - N(u - \vartheta)$ and putting $\Delta(u) = u^{-\frac{1}{2}}[N(u + \vartheta) - N(u - \vartheta)]$, we have

$$u^{-\frac{1}{2}}\gamma(u) > -\Delta(u)\left\{- \log \Delta(u) + \log\left(1 + \frac{\vartheta}{u}\right) - \log \frac{c}{2e}\right\} + O[u^{-\frac{1}{2}} \log u].$$

The right side tends to zero as $u \to \infty$ since $\Delta(u) \to 0$ by Hypothesis A. It follows that $\liminf u^{-\frac{1}{2}}\gamma(u) \geqq 0$, and the lemma is proved.

Lemma 5.3.[22] *If* $\{a_n\}$ *satisfies either* (i) *Hypothesis* C *or* (ii) *Hypothesis* B *with the added assumption that* $\inf (a_{n+1} - a_n) > 0$, *then*

$$\lim_{n\to\infty} a_n^{-\frac{1}{2}} \log | G'(a_n) | = 0.$$

Proof. We have

$$G'(a_n) = -\prod_{k\neq n} \left\{1 - \frac{a_n}{a_k}\right\}. \tag{5.2.9}$$

Let us now introduce the family of entire functions

$$G_n(w) = \prod_{k\neq n} \left\{1 - \frac{w}{a_k}\right\}, \tag{5.2.10}$$

and let $N_n(t)$ be the enumerative function of the sequence $a_1, a_2, \cdots, a_{n-1}, a_{n+1}, \cdots$. To these functions correspond functions $\omega_n(t)$, $\vartheta_n(t)$ and open sets $I_h(G_n)$. Since two functions $N_m(t)$ and $N_n(t)$ differ by at most one unit for any value of t and ultimately are equal to $N(t) - 1$, where $N(t)$ is the enumerative function of the sequence $\{a_n\}$, we can clearly replace $N_n(t)$, $\omega_n(t)$ and $\vartheta_n(t)$ in all estimates by $N(t)$, $\omega(t)$, and $\vartheta(t)$ without introducing any errors. It follows that given any $\delta > 0$ we can find a $U = U(\delta, h)$ such that

$$u^{-\frac{1}{2}} \log | G_n(u) | \geqq -\delta \quad \text{for} \quad u \geqq U(\delta, h), \qquad u \in I_h(G_n).$$

But $a_n \in I_h(G_n)$ for all large n and sufficiently small h. It follows that

$$a_n^{-\frac{1}{2}} \log | G_n(a_n) | = a_n^{-\frac{1}{2}} \log | G'(a_n) | \geqq -\delta$$

for $n \geqq n(\delta)$, whence the lemma follows.

In passing we note[23]

Lemma 5.4. *If* $G_n(w)$ *is defined by* (5.2.10) *and*

$$G_n(w) = \sum_{j=0}^{\infty} g_{n,j} w^j, \tag{5.2.11}$$

then for every given $\delta > 0$ *there exists a finite* $B(\delta)$ *independent of* j *and* n *such that*

$$| g_{n,j} | \leqq B(\delta) \frac{\delta^j}{(2j)!}. \tag{5.2.12}$$

[22] For the analogue of Hypothesis C, cf. Carlson and Landau [2], pp. 185-186.

[23] Cf. Carlson and Landau [2], p. 185.

The proof is an immediate consequence of Cauchy's estimates plus the fact that the maximum modulus of $G_n(w)$ is $G_n(-r) < G(-r)$.

We have next

LEMMA 5.5. *If* $\{a_n\}$ *satisfies Hypothesis* C *or Hypothesis* B *plus the assumption that* $\inf (a_{n+1} - a_n) > 0$, *and if*

$$H(w) = G(w) \sum_{n=1}^{\infty} \frac{B_n}{w - a_n}, \tag{5.2.13}$$

where $|B_n| \leqq B$ *for all* n *and*

$$\lim_{n \to \infty} a_n^{-\frac{1}{2}} \log |B_n| = 0, \tag{5.2.14}$$

then $H(w)$ *is an entire function whose order does not exceed that of* $G(w)$ *and which is of minimal type if the order equals* $\frac{1}{2}$. *Further*

$$\lim_{n \to \infty} a_n^{-\frac{1}{2}} \log |H(a_n)| = 0. \tag{5.2.14}$$

Proof. Suppose that $\inf (a_{n+1} - a_n) = 2h > 0$ and surround each zero a_n of $G(w)$ by a circle $C_n : |w - a_n| = h$. Let D be the domain outside of all circles C_n. For $|w| = r$, $w \in D$ we have

$$\sum_{n=1}^{\infty} \left| \frac{B_n}{w - a_n} \right| \leqq B\{\textstyle\sum' |w - a_n|^{-1} + \sum'' |r - a_n|^{-1}\},$$

where the first sum extends over all a_n for which $r_1 = r - r^{\frac{1}{2}} \leqq a_n \leqq r + r^{\frac{1}{2}} = r_2$ and the second sum takes in all the rest. The second sum equals

$$\left\{\int_0^{r_1} + \int_{r_2}^{\infty}\right\} \frac{dN(t)}{|r - t|} \leqq r^{-\frac{1}{2}} \int_0^{r_1} dN(t) + 2r^{\frac{1}{2}} \int_{r_2}^{\infty} \frac{dN(t)}{t}$$
$$\leqq r^{-\frac{1}{2}} N(r) + 2r^{\frac{1}{2}} \int_r^{\infty} \frac{N(t)}{t^2}\, dt \leqq 5[\omega^*(r)]^{-1} \to 0$$

as $r \to \infty$. In the case of Hypothesis C, the first sum contains at most a bounded number of terms and is consequently $O(1/h)$ uniformly in r. If Hypothesis B holds instead, the first sum need no longer be uniformly bounded, but it is at most $O(\log r)$, since $|w - a_n| > (j + 1)h$ if a_n is the j-th zero counting from the circle $|w| = r$ along the real axis either towards the origin or towards infinity and the number of terms in the sum is at most $O(r^{\frac{1}{2}})$. Consequently there exists a constant $M = M(h)$ such that in D

$$|H(w)| < M(h)G(-r)[1 + \log (r + 1)].$$

This estimate holds also on the circles C_n. Since $H(w)$ is holomorphic in C_n, we must have

$$|H(w)| < M(h)G(-r - 2h)[1 + \log (r + 1 + 2h)]$$

in C_n and hence everywhere in the plane. This estimate proves the statements about the order and type of $H(w)$ made in the theorem.

Since $H(a_n) = B_n G'(a_n)$, formula (5.2.15) follows from (5.2.14) combined with Lemma 5.3.

We now introduce a fourth assumption.[24]

HYPOTHESIS D. *Hypothesis* A *holds and there exists a* $\mu \geqq 0$ *such that*

$$a_{n+1} - a_n \geqq a_n^{\frac{1}{2}} \exp\left[-(\mu + \epsilon)\omega^*(n)\right], \qquad n \geqq n(\epsilon). \tag{5.2.16}$$

We have then

LEMMA 5.6. *If* $\{a_n\}$ *satisfies Hypothesis* D, *then*

$$\liminf u^{-\frac{1}{2}} \log |G(u)| \geqq -\mu \tag{5.2.17}$$

when $u \to \infty$ *in* I_h *and*

$$\liminf_{n\to\infty} a_n^{-\frac{1}{2}} \log |G'(a_n)| \geqq -\mu. \tag{5.2.18}$$

Proof. By virtue of Lemma 5.1 it is enough to prove that $\liminf u^{-\frac{1}{2}}\gamma(u) \geqq -\mu$ as $u \to \infty$ in I_h. The proof is analogous to that of Lemma 5.2 under Hypothesis C to which we refer the reader for notation and omitted details. For $n < l$ we get

$$u - a_n > (l - n)a_n^{\frac{1}{2}} \exp\left[-(\mu + \epsilon)\omega^*(n)\right]$$

and for $n > l + 1$

$$a_n - u > (n - l - 1)u^{\frac{1}{2}} \exp\left[-(\mu + \epsilon)\omega^*(u)\right].$$

Hence

$$\gamma(u) > 2 \log h + \log\left[(l-k)!\,(m-1-l)!\right] + \frac{1}{2}\sum_k^{l-1} \log a_n$$

$$+ (m-1-l)\log u - \sum_k^m \log a_n - (\mu+\epsilon)\sum_k^{l-1}\omega^*(n) - (\mu+\epsilon)(m-1-l)\omega^*(u).$$

By virtue of the proof of Lemma 5.2 we get

$$\liminf u^{-\frac{1}{2}}\gamma(u) \geqq -(\mu + \epsilon) \limsup\left[(m - 1 - k)u^{-\frac{1}{2}}\omega^*(u)\right].$$

But $m - 1 - k < m = N(u + \vartheta(u))$ and $\omega^*(u) \leqq \omega(u + \vartheta(u))$. Hence the last bracket is dominated by $[1 + \vartheta(u)/u]^{\frac{1}{2}} \to 1$. Thus $\liminf u^{-\frac{1}{2}}\gamma(u) \geqq -(\mu + \epsilon)$ for every $\epsilon > 0$ and (5.2.17) is proved. We obtain (5.2.18) from (5.2.17) by the same type of argument as was used in proving Lemma 5.3 with the aid of Lemma 5.2. This completes the proof of Lemma 5.6.

Our last lemma is the following

LEMMA 5.7. *If* $U(x)$ *is a continuous monotone increasing function,* $U(0) = 0$, *and* $U(x) = o(x^{\frac{1}{2}})$ *as* $x \to \infty$, *then there exists an entire function* $J(w)$ *of order* $\frac{1}{2}$ *and minimal type, real on the real axis and having only real negative zeros, such that* $J(x)\exp[-U(x)] \to \infty$ *as* $x \to \infty$.

[24] Cf. Carlson and Landau [2], pp. 187–188 for the analogues of Hypothesis D and Lemma 5.6.

Proof. We assume that $x^{-\rho}U(x) \to \infty$ for every $\rho < \frac{1}{2}$ since otherwise the construction is trivial. We can further suppose that $U(\alpha x)/U(x)$ is large for large values of α, uniformly in $x \geqq 1$. We can then find a constant a such that $U(ax) > 10\,U(x)$ for $x \geqq 1$. If $y = U(x)$, let $x = V(y)$ be the inverse function which is also monotone increasing and tends faster to infinity than any constant multiple of y^2 but slower than $y^{2+\epsilon}$, $\epsilon > 0$. Let us now form

$$J(w) = \prod_{n=1}^{\infty}\left\{1 + \frac{2aw}{V(n)}\right\}. \tag{5.2.19}$$

This is clearly an entire function of order $\frac{1}{2}$ and minimal type.

For $x > 0$ we have

$$\frac{J'(x)}{J(x)} = \sum_{n=1}^{\infty}\frac{2a}{2ax + V(n)} > 2a\int_1^{\infty}\frac{dt}{2ax + V(t)} = 2a\int_{V(1)}^{\infty}\frac{dU(s)}{2ax + s}$$

$$= -\frac{2a}{2ax + V(1)} + 2a\int_{V(1)}^{\infty}\frac{U(s)\,ds}{[2ax + s]^2}$$

and

$$2a\int_{2ax}^{\infty}\frac{U(s)\,ds}{[2ax + s]^2} > \frac{U(2ax)}{2x}.$$

It follows that for sufficiently large values of x

$$\frac{J'(x)}{J(x)} > \frac{U(2ax)}{3x}.$$

Since

$$\int_{\frac{1}{2}x}^{x} U(2at)\,\frac{dt}{t} > \log 2\; U(ax),$$

we conclude that ultimately

$$\log J(x) > \tfrac{1}{3}\log 2\; U(ax) > 2U(x),$$

and this inequality implies that $J(w)$ has the required properties.

5.3. **Absence of dispersion causing singularities.** We now take up the discussion of singularities of functions defined by Hermitian series and start with an analogue of the well-known Pringsheim-Vivanti theorem for power series.[25]

THEOREM 5.1. *If $i^n f_n \geqq 0$ for all large values of n and if $\tau > 0$ is the ordinate of convergence of the Hermitian series*

$$f(z) = \sum_{n=0}^{\infty} f_n h_n(z), \tag{5.3.1}$$

then $z = \tau i$ is a singular point of $f(z)$, whereas $z = -\tau i$ is singular if instead $(-i)^n f_n \geqq 0$.

[25] A. Pringsheim [20], p. 42, and G. Vivanti [27], p. 112.

Proof. It is enough to prove the first case. Assuming the conditions on the coefficients to hold for all n, as we may do without restricting the generality of the discussion, we have

$$
\begin{aligned}
F(z) \equiv e^{-\frac{1}{2}z^2} f(iz) &= \sum_0^\infty f_n H_n(iz) \\
&= \sum_{n=0}^\infty |f_n| \left\{ \sum_{2k \leq n} \binom{n}{2k} \frac{(2k)!}{k!} (2z)^{n-2k} \right\}.
\end{aligned}
\tag{5.3.2}
$$

This double series, if summed first for k and then for n, is convergent in the strip $-\tau < x < \tau$. The double series is absolutely convergent when $|z| < \tau$, since all coefficients are positive, and for such values of z the double series can be rearranged as a power series in z with positive coefficients. If the radius of convergence of this power series be denoted by R, then $z = R$ is a singular point of $F(z)$ by the classical Pringsheim-Vivanti theorem, i.e., $z = Ri$ is a singular point of $f(z)$. We have clearly $R \geqq \tau$. But if $R > \tau$, then the series

$$
F(z) = \sum_{m=0}^\infty \left\{ \sum_{k=0}^\infty \binom{m+2k}{2k} \frac{(2k)!}{k!} |f_{m+2k}| \right\} (2z)^m
$$

would converge for some real positive values of z greater than τ. All coefficients being positive, the series could be rearranged by interchange of the order of summation and we would conclude that formula (5.3.2) should be valid for a real value of z greater than τ. This is a contradiction, whence it follows that $R = \tau$ and that $z = \tau i$ is a singular point of $f(z)$.

As a corollary of this theorem we obtain

THEOREM 5.2. *If $f_{2k+1} = 0$ and $(-1)^k f_{2k} \geqq 0$ for all large values of k, then $f(z)$ has $z = \pm \tau i$ as singular points. The same conclusion is valid if instead $f_{2k} = 0$ and $(-1)^k f_{2k+1} \geqq 0$ for all large values of k.*

We prove next an analogue of a theorem due to O. Szász.[26] In order to state this theorem in a simple form, let us introduce the following terminology. We set $f_n = a_n + ib_n$ and define

$$
\begin{aligned}
f_{re}(z) &= \sum_0^\infty a_{2k} h_{2k}(z), & f_{ie}(z) &= \sum_0^\infty b_{2k} h_{2k}(z), \\
f_{ro}(z) &= \sum_0^\infty a_{2k+1} h_{2k+1}(z), & f_{io}(z) &= \sum_0^\infty b_{2k+1} h_{2k+1}(z).
\end{aligned}
\tag{5.3.3}
$$

We refer to these series as *the components of* $f(z)$ and have then

THEOREM 5.3. *If one of the components of $f(z)$ is singular at $z = \pm \tau i$, then $f(z)$ is singular at least at one of these points.*

Proof. It is enough to prove that the assumption that $f(z)$ is holomorphic at

[26] Loc. cit., Hilfsatz 1, pp. 100–101.

both points implies that all four components are holomorphic at these points. If $f(z)$ is holomorphic at $z = \tau i$,

$$f(z) = \sum_0^\infty (\alpha_n + i\beta_n)(z - \tau i)^n,$$

convergent for $|z - \tau i| < \rho$, say. On the imaginary axis for $\tau - \rho < y < \tau + \rho$

$$\sum_0^\infty \alpha_{2k}(z - \tau i)^{2k} + i \sum_0^\infty \beta_{2k+1}(z - \tau i)^{2k+1} \tag{5.3.4}$$

is the real component of $f(z)$, whereas

$$\sum_0^\infty \beta_{2k}(z - \tau i)^{2k} - i \sum_0^\infty \alpha_{2k+1}(z - \tau i)^{2k+1} \tag{5.3.5}$$

is the imaginary one. On the other hand,

$$f(z) = [f_{re}(z) + if_{i0}(z)] + i[f_{ie}(z) - if_{r0}(z)]$$

and here each bracket is real on the imaginary axis for $-\tau < y < \tau$. It follows that the first bracket coincides with the series in (5.3.4) and the second one with the series in (5.3.5) on the common interval of convergence. In other words,

$$f_{re}(z) + if_{i0}(z) \quad \text{and} \quad f_{ie}(z) - if_{r0}(z)$$

are both holomorphic at $z = \tau i$. Using the parity properties of the components, we conclude that

$$f_{re}(z) - if_{i0}(z) \quad \text{and} \quad f_{ie}(z) + if_{r0}(z)$$

are holomorphic at $z = -\tau i$.

But $f(z)$ was also supposed to be holomorphic at $z = -\tau i$. Proceeding in the same manner at this point, we see that

$$f_{re}(z) + if_{i0}(z) \quad \text{and} \quad f_{ie}(z) - if_{r0}(z)$$

are holomorphic at $z = -\tau i$ and by parity considerations

$$f_{re}(z) - if_{i0}(z) \quad \text{and} \quad f_{ie}(z) + if_{r0}(z)$$

are holomorphic at $z = \tau i$. Adding and subtracting we see that all four components are holomorphic at $z = \pm\tau i$. This completes the proof of the theorem.

Combining the last two theorems, we get

THEOREM 5.4. *If any one of the component series of $f(z)$ satisfies the conditions of Theorem 5.1, and if the ordinate of convergence of this series also equals τ, then $f(z)$ cannot be holomorphic both at $z = \tau i$ and at $z = -\tau i$.*

A particular case of this theorem is the following analogue of the theorem of P. Dienes [5] for power series.

THEOREM 5.5. *If there exist two angles, θ_1 and θ_2, where $0 < \theta_2 - \theta_1 < \pi$, such that $\theta_1 \leqq \arg (i^n f_n) \leqq \theta_2$ for all large values of n, then $z = \tau i$ is a singular point of $f(z)$. If the condition is merely satisfied for all large even values of n, and if the terms of even order have the same ordinate of convergence as the whole series, then at least one of the points $\pm \tau i$ is singular. The same conclusion is valid if the condition is satisfied by the odd terms instead.*

It should be observed that the first statement of this theorem is not a direct consequence of Theorem 5.4, which would not give quite so specific information. It follows, however, from the proof of Theorem 5.1 applied to the present case plus the original theorem of Dienes on power series.

5.4. **Slow dispersion and singularities.** We shall now show that dispersion of the arguments of the coefficients, if sufficiently slow, is not detrimental to the existence of singularities at the points $\pm\tau i$. If we are satisfied with information to the effect that at least one of these points is singular, we can restrict ourselves to finding conditions under which one of the component series will be singular at $\pm\tau i$.

Let $\{c_n\}$ be a given sequence of real numbers ($n = 0, 1, 2, \cdots$) of which c_0 can be assumed non-negative without restricting the generality. Suppose that c_n keeps a constant sign or is zero for $n_k \leqq n \leqq n_{k+1} - 1$ and c_n is different from zero for $n = n_k$, its sign for $n = n_k$ being $(-1)^k$. This defines a sequence of integers $\{n_k\}$. With this sequence we form the functions $N(t)$, $\omega(t)$, $\omega^*(t)$, and $G(w)$ as in §5.2. We then have[27]

THEOREM 5.6. *If the sequence $\{n_k\}$ satisfies either of the Hypotheses* B, C, *or* D *(with $\mu = 0$) of §5.2, then $z = \pm\tau i$ are singular points of the series*

$$\sum_{n=0}^{\infty} (-1)^n c_n h_{2n}(z) \quad \text{and} \quad \sum_{n=0}^{\infty} (-1)^n c_n h_{2n+1}(z), \tag{5.4.1}$$

$\tau > 0$ being the ordinate of convergence of the series.

Remark. The assumptions imply that the integers n_k where the sign changes take place satisfy one of the following assumptions:

(B) $n_k(k \log k)^{-2} \to \infty$,

(C) $n_k k^{-2} \to \infty$, $n_{k+1} - n_k > c n_k^{\frac{1}{2}}$,

(D) $n_k k^{-2} \to \infty$, $n_{k+1} - n_k > n_k^{\frac{1}{2}} \exp[-\epsilon\eta^*(k)]$, where $\eta^*(k) = \inf_{m \geqq k} n_m m^{-2}$.

Proof. It is enough to give the proof in the even case. We consider then the series

$$f(z) = \sum_{n=0}^{\infty} (-1)^n c_n h_{2n}(z) \tag{5.4.2}$$

[27] See O. Szász [23], Satz I, pp. 104–105 for a similar theorem on Dirichlet series based on the analogue of Hypothesis C.

and the transformed series

$$F(z) = G(\tfrac{1}{4}\delta_z)f(z) = \sum_{n=0}^{\infty} (-1)^n c_n G(n + \tfrac{1}{4}) h_{2n}(z), \tag{5.4.3}$$

where, as already mentioned,

$$G(w) = \prod_{k=1}^{\infty} \left\{1 - \frac{w}{n_k}\right\}.$$

This is an entire function of order $\rho \leqq \frac{1}{2}$ and of minimal type if $\rho = \frac{1}{2}$ and $G(w)$ satisfies the conditions of Lemma 5.2 in the case in which Hypothesis B or C holds and Lemma 5.6 in case of D (with $\mu = 0$). Clearly $n + \frac{1}{4}$ belongs to the set $I_h(G)$ for $h < \frac{1}{4}$ and every n. Hence

$$\lim_{n\to\infty} (4n + 1)^{-\frac{1}{2}} \log | G(n + \tfrac{1}{4}) | = 0$$

and the two series (5.4.2) and (5.4.3) have the same ordinate of convergence, τ say. Further, for $n_k \leqq n < n_{k+1}$ both c_n and $G(n + \frac{1}{4})$ have the sign $(-1)^k$, i.e., $c_n G(n + \frac{1}{4}) \geqq 0$. Theorem 5.2 then shows that $z = \pm\tau i$ are singular points of $F(z)$ and Theorem 4.2 informs us that $F(z)$ can be continued analytically along any finite path along which $f(z)$ can be so continued. Hence $z = \pm\tau i$ are singular points of $f(z)$, and the theorem is proved.

The interested reader can combine Theorems 5.3 and 5.6 to get further results concerning slow dispersion. We shall pass over to our main result on singularities.

5.5. A gap theorem.

We shall prove[28]

THEOREM 5.7. *Let $\{n_k\}$ be a sequence of integers satisfying either of Hypotheses* B, C, *or* D (*with* $\mu = 0$). *Then the function $f(z)$ defined by*

$$f(z) = \sum_{k=1}^{\infty} f_{n_k} h_{n_k}(z) \tag{5.5.1}$$

has the strip of convergence of the series as its natural domain of existence.

Proof. Let us consider an arbitrary point $x_0 + i\tau$ on the upper line of convergence. We assume $0 < \tau < \infty$. By Theorem 1.4

$$h_n(x_0 + iy) = e^{-i\nu x_0} h_n(iy)\left\{1 + O\left(\frac{1}{\nu}\right)\right\}, \qquad \nu = (2n + 1)^{\frac{1}{2}}, \tag{5.5.2}$$

[28] This is the Fabry theorem. It has been proved for Dirichlet series under the analogue of Hypothesis C by Szász [23], Satz III, p. 106, and Carlson and Landau [2], Satz I, p. 185, and under the analogue of Hypothesis D with $\mu = 0$ by Carlson and Landau [2], Satz II, p. 187. The main idea of the proof below, i.e., the construction of a transformed series the sum of which tends to infinity for vertical approach to a given point on one of the lines of convergence, seems to be new in this connection, though it is of course closely related to classical devices.

where the estimate holds uniformly for $\frac{1}{2}\tau < y < \tau$. Let us put

(5.5.3) $$f_n e^{-inx_0} = f_n(x_0), \qquad n = n_k.$$

We consider the sequence $\{a_k\}$ where $a_k = 2n_k + 1$ and construct the corresponding entire function

$$G(w) = \prod_{k=1}^{\infty}\left(1 - \frac{w}{a_k}\right).$$

It is clear that the sequence $\{a_k\}$ satisfies the same Hypothesis B, C, or D as $\{n_k\}$. Using $G(w)$ and taking

(5.5.4) $$B_k = e^{i\beta_k}, \qquad \beta_k = -(2k + n_k)\tfrac{1}{2}\pi - \arg[f_{n_k}(x_0)],$$

we construct the function $H(w)$ of Lemma 5.5. These choices are evidently consistent with the conditions of Lemma 5.5. We note that $H(a_k) = G'(a_k)e^{i\beta_k}$ and $\operatorname{sgn} G'(a_k) = (-1)^k$.

Let us now form the transformed series

(5.5.5) $$F_0(z) = H(\delta_z)f(z) = \sum_{k=1}^{\infty} H(a_k)f_{n_k}h_{n_k}(z).$$

In view of (5.2.15) the ordinate of convergence of this series also equals τ and by Theorem 4.2 the point $z = x_0 + i\tau$ is a singular point of $F_0(z)$ only if it is a singular point of $f(z)$. In order to prove that it is actually a singular point of $F_0(z)$, we consider first the auxiliary series

(5.5.6) $$\Phi(z) = \sum_{k=1}^{\infty} H(a_k)f_{n_k}(x_0)h_{n_k}(z).$$

This series is obtained from (5.5.5) by expressing $h_{n_k}(x_0 + iy)$ with the aid of (5.5.2), dropping all remainder terms, and replacing iy by z in the result. In view of formula (5.5.4) we have $i^{n_k}H(a_k)f_{n_k}(x_0) > 0$. It follows that $\Phi(z)$ has a singular point at $z = \tau i$.

It is then fairly plausible that $z = x_0 + i\tau$ shall be a singular point of $F_0(z)$. This will certainly be the case if the series in (5.5.6) diverges for $z = \tau i$. Indeed, all the terms of the series are positive on the imaginary axis, and if the series diverges at $z = \tau i$, this implies that $\Phi(iy) \to +\infty$ as y increases to the value τ. On the other hand, the ratio between corresponding terms in the two series for $\Phi(iy)$ and $F_0(x_0 + iy)$ tends to one as the subscript tends to infinity. From this we can conclude that $\Re[F_0(x_0 + iy)] \to +\infty$ as y increases to τ, whence it follows that $z = x_0 + i\tau$ is a singular point of $F_0(z)$ and consequently also of $f(z)$. Since the absolute values of the coefficients in (5.5.6) are independent of the particular value of x_0, it follows that this argument applies either at all points of the upper line of convergence $y = \tau$ or at no point. In the former case the argument also applies at all points of the lower line of convergence, $y = -\tau$, if in the definition of $H(w)$ we replace each coefficient B_k by its conjugate. Thus,

in the divergency case the proof of the theorem is complete. This case is characterized simply by the divergence of the series

$$(5.5.7)\qquad \sum g_k \equiv \sum |G'(a_k)|\,|f_{n_k}|\,A_{n_k} \exp(\tau a_k^{\frac{1}{2}}), \qquad a_k = 2n_k + 1.$$

In case this series does not diverge, we have to modify the argument. We observe first that it is permissible to assume that

$$(5.5.8)\qquad g_k \geqq \exp(-\epsilon a_k^{\frac{1}{2}}), \qquad k > k(\epsilon),$$

no matter how small $\epsilon > 0$ is. Indeed, if there should exist an infinite subsequence $\{g_{k_j}\}$ tending faster to zero than is permitted by (5.5.8), then the corresponding terms of (5.5.1) form an infinite series having an ordinate of convergence greater than τ. The function $f(z)$ is then the sum of two Hermitian series, one having the ordinate of convergence equal to τ, the other greater than τ. It is no restriction of the generality to assume that the second series has been suppressed and that consequently (5.5.8) is valid.

In an x,y-plane let us mark the points $P_0 = (0, 0)$ and $P_k = (a_k, -\log g_k)$ and draw a polygonal line of least ordinate, concave downwards, such that all the points P_k are either on or below the line. Let $y = U(x)$ be the equation of this line. It is clear that $U(x)$ is continuous, monotone increasing, and $o(x^{\frac{1}{2}})$ as $x \to \infty$. To this function $U(x)$ we construct the entire function $J(w)$ of Lemma 5.7 and with the aid of this function we form the transformed series

$$(5.5.9)\qquad J(\delta_z)f(z) = j(z) = \sum_{k=1}^{\infty} J(a_k) f_{n_k} h_{n_k}(z).$$

This series also has the ordinate of convergence τ and $j(z)$ has no other singularities in the finite plane than those of $f(z)$. To this new series we can apply the operator $H(\delta_z)$, where $H(w)$ is determined as above, and we are now sure to be in the divergency case because the critical series (5.5.7) is now

$$(5.5.10)\qquad \sum g_k^* = \sum |G'(a_k)|\,|J(a_k)|\,|f_{n_k}|\,A_{n_k} \exp(\tau a_k^{\frac{1}{2}}),$$

and the properties of $J(w)$ show that the terms of this series tend to infinity. The previous analysis then applies and shows that the lines of convergence, $y = \pm\tau$, are singular lines of $j(z)$ and consequently also of $f(z)$. This completes the proof of the theorem. An interesting complement of Theorem 5.7 is given in §6.4.

5.6. **A gap theorem with limited continuability.** The preceding theorem states that long and fairly regular gaps imply that a Hermitian series has its domain of convergence as its domain of existence. If the restrictions imposed upon the subscripts $\{n_k\}$ are made less severe, we naturally expect the conclusion to cease holding, but we may still be able to prove that the domain of existence of the function defined by the series is definitely limited. The following analogue of a theorem proved by F. Carlson and E. Landau [2], p. 187 for Dirichlet series is an example of such a result.

THEOREM 5.8. *Let $\{n_k\}$ be a monotone increasing sequence of integers satisfying Hypothesis* D *of* §5.2. *Let τ, $0 < \tau < \infty$, be the ordinate of convergence of the series*

$$f(z) = \sum_{k=1}^{\infty} f_{n_k} h_{n_k}(z). \tag{5.6.1}$$

Then $f(z)$ cannot be continued analytically across either of the lines $y = \pm(\tau + \mu)$, i.e., the domain of existence of $f(z)$ is a subset of the strip $-(\tau + \mu) < y < \tau + \mu$.

Proof. We can imitate the procedure followed by Carlson and Landau. As in §5.5 we take $a_k = 2n_k + 1$ and form the corresponding entire function $G(w)$. We also need the entire functions $G_n(w)$ of Lemmas 5.3 and 5.4. We start by forming the transformed series

$$-G'(\delta_z)f(z) = F(z) = -\sum_{k=1}^{\infty} G'(a_k) f_{n_k} h_{n_k}(z). \tag{5.6.2}$$

In this case the transformation has the effect of both changing the ordinate of convergence and removing singularities, but we can still get useful information from the transformed series.

Suppose that it should be possible to continue $f(z)$ analytically along a finite path from $z = 0$ to $z = z_0$, where z_0 is located on one of the critical lines $y = \pm(\tau + \mu)$, and that $f(z)$ is holomorphic in a circle $|z - z_0| < 4\rho$, say. Let $z_1 = z_0 \pm \rho i$, where the sign is the same as that of the imaginary part of z_0. By virtue of formula (3.3.12) there exists a finite quantity $B_1(\rho)$ such that

$$|\delta_z^j f(z_1)| \leqq B_1(\rho)(2j)!(2\rho)^{-2j} \tag{5.6.3}$$

for all j. On the other hand, $G'(w)$ being an entire function of order $\frac{1}{2}$ and minimal type at most, Theorem 3.2 applies and shows that

$$-G'(a_k) f_{n_k} h_{n_k}(z) = G_k(a_k) f_{n_k} h_{n_k}(z) = G_k(\delta_z) f(z) = \sum_{j=0}^{\infty} g_{k,j} \delta_z^j f(z),$$

the infinite series being convergent in any domain where $f(z)$ is holomorphic. In particular we have convergence at $z = z_0 \pm \rho i$, and using Lemma 5.4 we see that

$$|G'(a_k) f_{n_k} h_{n_k}(z_1)| \leqq B(\delta) B_1(\rho) \sum_{j=0}^{\infty} \delta^j (2\rho)^{-2j}.$$

This expression is bounded for all k if we take $\delta = \rho^2$ as is permissible.

But this result implies that the terms of the series (5.6.2) are bounded at a point z whose distance from the real axis is $\tau + \mu + \rho$. Theorem 2.1 then shows that the series has an ordinate of convergence $\tau_1 \geqq \tau + \mu + \rho$. But in view of Lemma 5.6, the ordinate of convergence of the series

$$f(z) = \sum_{k=1}^{\infty} G'(a_k)[G'(a_k)]^{-1} f_{n_k} h_{n_k}(z)$$

is at least $\tau_1 - \mu$. We have consequently $\tau + \mu + \rho \leqq \tau_1 \leqq \tau + \mu$, and this is a contradiction for any positive value of ρ. Hence it is impossible for $f(z)$ to be holomorphic anywhere on the lines $y = \pm(\tau + \mu)$, and the theorem is proved.

In connection with this gap theorem it would be interesting to investigate if the phenomenon of *over-convergence*, the fundamental importance of which has been made manifest by A. Ostrowski in the theory of power series and Dirichlet series, does not also play a decisive rôle in the theory of Hermitian gap-series. Some results in this connection can be read off from Ostrowski's results,[29] but a closer investigation has to be postponed to a later occasion.

5.7. **Irregular dispersion and non-continuability.** It is well known that a power series is non-continuable if the arguments of the coefficients are dispersed at random. One expression of this fact is a famous theorem of Fatou which was completely proved by A. Hurwitz and G. Pólya.[30] The argument of Hurwitz applied to Hermitian series gives

THEOREM 5.9. *Let*

$$f(z) = \sum_{n=0}^{\infty} f_n h_n(z) \tag{5.7.1}$$

be a given Hermitian series having the ordinate of convergence τ, $0 < \tau < \infty$. *Let* $\mathfrak{H}$ *be the class of all Hermitian series of the form*

$$f_\epsilon(z) = \sum_{n=0}^{\infty} \epsilon_n f_n h_n(z), \tag{5.7.2}$$

where $\epsilon_n = \pm 1$. *The set of series in* $\mathfrak{H}$ *which cannot be continued analytically outside of the strip of convergence is non-denumerable.*

Proof. We begin by picking out a subsequence $\{n_k\}$ subject to the conditions
(i) $n_k(k \log k)^{-2} \to \infty$,
(ii) $\lim\limits_{k\to\infty} (2n_k + 1)^{-\frac{1}{2}} \log (A_{n_k} |f_{n_k}|) = -\tau$,
and put

$$\varphi(z) = \sum_{k=1}^{\infty} f_{n_k} h_{n_k}(z). \tag{5.7.3}$$

This series has τ as its ordinate of convergence and cannot be continued outside of its strip of convergence by virtue of Theorem 5.7. Let

$$\varphi(z) = \sum_{j=1}^{\infty} \varphi_j(z), \tag{5.7.4}$$

[29] See in particular A. Ostrowski [16], Theorem 3, pp. 333-334 and gap theorem, p. 335. Ostrowski is mainly concerned here with sequences of polynomials which converge as the partial sums of a Taylor series, i.e., where the remainder after n terms is $O(r^n)$ with $r < 1$. In the case of expansions in terms of Hermite or Laguerre polynomials, n has to be replaced by $n^{\frac{1}{2}}$ in the remainder. The general methods of Ostrowski seem to apply, however, also to such expansions.

[30] See P. Fatou [10], p. 400, and Hurwitz-Pólya [13].

where each function $\varphi_j(z)$ is a series of the form

$$\varphi_j(z) = \sum_{k=1}^{\infty} f_{m_k} h_{m_k}(z), \tag{5.7.5}$$

the subscripts $\{m_k\}$ being an infinite subset of the set $\{n_k\}$. It is understood that each term $f_{n_k}h_{n_k}(z)$ of $\varphi(z)$ occurs in one and only one of the functions $\varphi_j(z)$. Thus every function $\varphi_j(z)$ also has the strip $-\tau < y < \tau$ as its domain of existence. Put

$$R(z) = f(z) - \varphi(z) \tag{5.7.6}$$

and define

$$F_\epsilon(z) = R(z) + \sum_{j=1}^{\infty} \epsilon_j \varphi_j(z), \tag{5.7.7}$$

where $\epsilon_j = \pm 1$. It is clear that every function $F_\epsilon(z)$ belongs to the set $\mathfrak{H}$. The functions $\{F_\epsilon(z)\}$ form a non-denumerable subset $\mathfrak{H}_1$ of $\mathfrak{H}$ and we shall prove that the subset of $\mathfrak{H}_1$ formed by series continuable outside of the strip of convergence is at most denumerable. A series $F_\epsilon(z)$ which can be continued analytically outside of the strip of convergence is holomorphic in a certain set of intervals on the lines of convergence, I_ϵ say. The sets I_ϵ and I_η corresponding to two different sequences of signs, $\{\epsilon_j\}$ and $\{\eta_j\}$, cannot overlap. Indeed, if they did, then

$$F_\epsilon(z) - F_\eta(z) = \sum_{j=1}^{\infty} (\epsilon_j - \eta_j)\varphi_j(z)$$

would be holomorphic in at least one interval on one of the lines of convergence. Because of the choice of the sequence $\{n_k\}$, this contradicts Theorem 5.7. The set $\{I_\epsilon\}$, consisting of non-overlapping systems of intervals, can be at most denumerable. Hence the subset of continuable Hermitian series of the type (5.7.7) is at most denumerable. Consequently, the complementary subset of $\mathfrak{H}_1$, which is made up of non-continuable Hermitian series of type (5.7.7), is non-denumerable and the theorem is proved.

Chapter 6. The associated Dirichlet and Fourier series

6.1. **Definitions.** We shall make a systematic study of *the associated series of a given Hermitian series* which made their first appearance in Chapter 2.

With a given Hermitian series

$$f(z) = \sum_{n=0}^{\infty} f_n h_n(z) \tag{6.1.1}$$

we associated the Fourier series

$$C(z; f) = C(z) = \sum_{n=0}^{\infty} f_n c_n(z), \tag{6.1.2}$$

$$S(z;f) = S(z) = \sum_{n=0}^{\infty} f_n s_n(z), \tag{6.1.3}$$

where $c_n(z)$ and $s_n(z)$ are the functions defined in formulas (1.2.4) and (1.2.5). It was shown in Theorem 2.3 that the first two series are simultaneously convergent or divergent and the same argument applies also to the third series.

We also had occasion to use the associated Dirichlet series

$$E^+(z;f) = E^+(z) = \sum_{n=0}^{\infty} f_n e_n^+(z), \tag{6.1.4}$$

$$E^-(z;f) = E^-(z) = \sum_{n=0}^{\infty} f_n e_n^-(z), \tag{6.1.5}$$

where

$$e_n^+(z) = A_n(-i)^n \exp\left[i(2n+1)^{\frac{1}{2}}z\right], \qquad e_n^-(z) = A_n i^n \exp\left[-i(2n+1)^{\frac{1}{2}}z\right]. \tag{6.1.6}$$

We note the obvious relations

$$C(z;f) = \tfrac{1}{2}\left[E^+(z;f) + E^-(z;f)\right], \tag{6.1.7}$$

$$S(z;f) = \frac{1}{2i}\left[E^+(z;f) - E^-(z;f)\right]. \tag{6.1.8}$$

If $\tau > 0$ is the ordinate of convergence of the series (6.1.1), then the function $E^+(z;f)$ is holomorphic in the half plane $y > -\tau$, whereas $E^-(z;f)$ is holomorphic in $y < \tau$. The above relations show that in the former half plane the functions $C(z;f)$ and $S(z;f)$ have no other singular points than those of $E^-(z;f)$ and these points are effectively singularities of the same nature of all three functions. Here it is assumed that the analytic continuation is carried out along finite paths restricted to the half plane in question. Similarly, the singularities of $C(z;f)$ and $S(z;f)$ in the half plane $y < \tau$ for analytic continuation restricted to this half plane are those of $E^+(z;f)$ which are effective singularities common to all three functions. In particular, *the Mittag-Leffler principal stars of the functions $C(z;f)$ and $S(z;f)$ are identical.*

6.2. **The expression of a Hermitian series in terms of its associated series.** Our point of departure is the basic integral equation of Chapter 1, viz.,

$$h_n(z) = c_n(z) + \frac{1}{\nu}\int_0^z t^2 \sin \nu(z-t)\, h_n(t)\, dt, \tag{6.2.1}$$

and the resulting expansion

$$\begin{aligned} h_n(z) &= c_n(z) \\ &+ \sum_{k=1}^{\infty} \int_0^z \int_0^{t_1} \cdots \int_0^{t_{k-1}} (t_1 t_2 \cdots t_k)^2 K(T_k, n) c_n(t_k)\, dt_k\, dt_{k-1} \cdots dt_1, \end{aligned} \tag{6.2.2}$$

where

$$K(T_k, n) = \nu^{-k} \prod_{j=1}^{k} \sin \nu(t_{j-1} - t_j), \qquad t_0 = z, \tag{6.2.3}$$

and $\nu = (2n + 1)^{\frac{1}{2}}$ as usual.

We can express $K(T_k, n)c_n(t_k)$ linearly in terms of the functions e_n^+ and e_n^- with suitable arguments. Let V_k be a k-vector whose component in the j-th place is $t_{j-1} - t_j$ ($j = 1, 2, \cdots, k$). Let U_k be any one of the 2^k k-vectors whose components are ± 1, and let (U_k, V_k) be the scalar product of these two vectors, i.e., the linear form $\sum_{j=1}^{k} \pm (t_{j-1} - t_j)$. We have then

$$\begin{aligned} K(T_k, n)c_n(t_k) &= \tfrac{1}{2}(2\nu i)^{-k} \\ &\cdot \sum (-1)^{\mu}\{e_n^+((U_k, V_k) + t_k) + e_n^-((U_k, V_k) - t_k)\}, \end{aligned} \tag{6.2.4}$$

where the summation extends over all k-vectors U_k and μ equals the number of negative components in U_k. Hence

$$\begin{aligned} h_n(z) = \tfrac{1}{2}[e_n^+(z) + e_n^-(z)] &+ \tfrac{1}{2} \sum_{k=1}^{\infty} (2\nu i)^{-k} \sum (-1)^{\mu} \int_0^z \int_0^{t_1} \cdots \int_0^{t_{k-1}} \\ \times (t_1 t_2 \cdots t_k)^2 [e_n^+((U_k, V_k) &+ t_k) + e_n^-((U_k, V_k) - t_k)]\, dt_k\, dt_{k-1} \cdots dt_1. \end{aligned} \tag{6.2.5}$$

Let us now introduce the functions

$$E_k^+(z; f) = E_k^+(z) = \sum_{n=0}^{\infty} \nu^{-k} f_n e_n^+(z), \tag{6.2.6}$$

$$E_k^-(z; f) = E_k^-(z) = \sum_{n=0}^{\infty} \nu^{-k} f_n e_n^-(z). \tag{6.2.7}$$

It is clear that these functions are k-fold integrals of $E^+(z)$ and $E^-(z)$, respectively. More precisely,

$$E_k^+(z) = i^k \frac{1}{k!} \int_{i\infty}^{z} (z - t)^{k-1} E^+(t)\, dt, \tag{6.2.8}$$

and a similar formula holds for $E_k^-(z)$ with i replaced by $-i$ both in the multiplier and in the lower limit of integration. Formula (6.2.6) shows that to every given $\epsilon > 0$ we can find an $M(\epsilon)$ such that $|E_k^+(z)| \leqq M(\epsilon)$ for all k and all z such that $y \geqq \epsilon - \tau$. Similarly $|E_k^-(z)| \leqq M(\epsilon)$ for all k and all z such that $y \leqq \tau - \epsilon$.

Suppose now that $E^+(z)$ can be continued analytically along a finite path C leading from $z = 0$ into the half plane $y < -\tau$. Formula (6.2.8) shows that for every k we can continue $E_k^+(z)$ along C as an analytic function. Moreover, if D is a simply-connected bounded domain containing C such that at all points of D the function $E^+(z)$ is holomorphic and less than a fixed constant in absolute value, then we can find an $M = M(D)$ such that $|E_k^+(z)| \leqq M(D)$ for all k and all z in D. This is an immediate consequence of formula (6.2.8). Similar

inequalities hold for $E_k^-(z)$ in any bounded simply-connected domain in which $E^-(z)$ is bounded and holomorphic.

After these preliminaries we can get the required representation of $f(z)$ in terms of $E^+(z; f)$ and $E^-(z; f)$. For this purpose we multiply both sides of formula (6.2.5) by f_n and sum for n, assuming to start with that z lies in the strip of convergence of the series (6.1.1). The result is

$$\begin{aligned} f(z) &= \tfrac{1}{2}[E^+(z) + E^-(z)] + \tfrac{1}{2}\sum_{k=1}^{\infty}(2i)^{-k}\sum(-1)^{\mu}\int_0^z\int_0^{t_1}\cdots\int_0^{t_{k-1}} \\ &\times (t_1t_2\cdots t_k)^2[E_k^+((U_k, V_k)+t_k) + E_k^-((U_k, V_k)-t_k)]\,dt_k\,dt_{k-1}\cdots dt_1. \end{aligned} \tag{6.2.9}$$

Our first task is to justify the process by means of which the formula was obtained and to prove that the series converges at least for $-\tau < y < \tau$. In order to prove this it is enough to show that if all terms in (6.2.5) are replaced by their absolute values and the resulting expressions are multiplied by $|f_n|$ and summed for n, we still get a convergent double series for $-\tau < y < \tau$. In (6.2.5) the paths of integration are arbitrary, but in (6.2.9) we have to restrict the paths so that $(U_k, V_k) + t_k$ and $(U_k, V_k) - t_k$ stay within the domains of definition of $E^+(w;f)$ and $E^-(w;f)$, respectively. The simplest way of attaining this object is to make all paths rectilinear. We have then

$$t_1 = r_1z, \quad t_2 = r_2t_1 = r_1r_2z, \quad \cdots, \quad t_k = r_1r_2\cdots r_kz, \tag{6.2.10}$$

where $0 \leqq r_j \leqq 1$. Consequently, putting $r_0 = 1$, we get

$$(U_k, V_k) \pm t_k = \left\{\sum_{j=0}^{k-1} \pm r_0r_1\cdots r_j(1 - r_{j+1}) \pm r_0r_1\cdots r_k\right\}z \equiv R_kz.$$

Here R_k is a real number lying between -1 and $+1$ and for a suitable choice of the r's and of the signs, R_k takes on any prescribed value in the interval $[-1, +1]$. But in any case

$$-|y| \leqq \Im[(U_k, V_k) \pm t_k] \leqq |y|,$$

and for all values of k and choices of U_k

$$\sum_{n=0}^{\infty} \nu^{-k}|f_n||e_n^+((U_k, V_k) + t_k)| \leqq \sum_{n=0}^{\infty} A_n|f_n|e^{\nu|y|} \equiv F(y),$$

a bounded quantity for $-(\tau - \epsilon) \leqq y \leqq \tau - \epsilon$. The same estimate holds for the terms involving $e_n^-(\cdot)$. Hence the double series in question is dominated by

$$F(y)\left\{1 + \sum_{k=1}^{\infty} r^{3k}\int_0^1\int_0^1\cdots\int_0^1 r_1^{3k-1}r_2^{3k-4}\cdots r_k^2\,dr_k\,dr_{k-1}\cdots dr_1\right\} = F(y)e^{\frac{1}{3}r^3}, \tag{6.2.11}$$

where $r = |z|$. The double series is consequently absolutely convergent and can be summed in any manner desired. Consequently formula (6.2.9) is valid at least in the strip $-\tau < y < \tau$. Naturally its true range of validity is the

domain of analyticity of the right side. In this connection we shall prove the following fundamental

THEOREM 6.1. *Let $\mathfrak{S}$ be the Mittag-Leffler principal star of $C(z; f)$, let $-\mathfrak{S}$ be the negative of this point set, and let $\mathfrak{C}$ be the cross section of $\mathfrak{S}$ and $-\mathfrak{S}$. Then $f(z)$ is holomorphic and representable by formula (6.2.9) in $\mathfrak{C}$.*

Remark. We can also characterize $\mathfrak{C}$ as the cross section of the Mittag-Leffler principal stars of the functions $E^+(z;f)$, $E^+(-z;f)$, $E^-(z;f)$, and $E^-(-z;f)$.

Proof. Suppose that $z_0 \in \mathfrak{C}$. This means that we can join z_0 with $-z_0$ by a line segment at all points of which $E^+(z; f)$ and $E^-(z; f)$ together with their integrals of all orders are holomorphic. We shall show that the right side of (6.2.9) is holomorphic at all points of this line segment. This is evidently true of the individual terms in the expansion. Indeed, in the k-fold integrals the expressions $(U_k, V_k) \pm t_k$ take on values on the line segment joining z with $-z$, no matter what permissible set of values we substitute for $t_1, t_2, \cdots, t_k$. Consequently all the functions $E_k^+((U_k, V_k) + t_k)$ and $E_k^-((U_k, V_k) - t_k)$ are holomorphic functions of z at all points of the line segment joining z_0 with $-z_0$. This property is evidently not going to be spoiled by multiplication by $(t_1t_2 \cdots t_k)^2$ and integration with respect to $t_1, t_2, \cdots, t_k$. If the reader prefers, he can make the change of variables of integration used above in (6.2.10). The discussion in terms of the new variables is still simpler. At any rate, the individual terms of (6.2.9) are holomorphic functions of z in $\mathfrak{C}$, so it is merely a question of convergence.

Let $\mathfrak{C}_0$ be a bounded domain, star-shaped and symmetric with respect to the origin, contained in $\mathfrak{C}$, and such that the distance of $\mathfrak{C}_0$ from the boundary of $\mathfrak{C}$ is positive. $\mathfrak{C}_0$ is a domain D of the type discussed in connection with formula (6.2.8). We can consequently find a constant $M = M(\mathfrak{C}_0)$ such that

$$|E_k^+((U_k, V_k) + t_k)| \leqq M(\mathfrak{C}_0), \qquad |E_k^-((U_k, V_k) - t_k)| \leqq M(\mathfrak{C}_0)$$

for all k, all choices of U_k, all values of z in $\mathfrak{C}_0$ and all corresponding permissible choices of $t_1, t_2, \cdots, t_k$. But then we see immediately that the series in (6.2.9) is dominated by an expression of type (6.2.11) with $F(y)$ replaced by $M(\mathfrak{C}_0)$. It follows that the series converges uniformly in $\mathfrak{C}_0$. The terms being holomorphic functions of z, we conclude that $f(z)$ is represented by the series (6.2.9) in $\mathfrak{C}$ and is holomorphic in $\mathfrak{C}$.

6.3. **Absence of singularities on the lines of convergence of a Hermitian series.** We shall derive several consequences of Theorem 6.1 and start with

THEOREM 6.2. *There are Hermitian series having a finite positive ordinate of convergence which have no finite singular points on the lines of convergence. In particular, there are entire functions having this property.*

Proof. It is enough to exhibit a cosine series

$$C(z) = \sum_{n=0}^{\infty} (-1)^n A_{2n} f_{2n} \cos (4n + 1)^{\frac{1}{2}} z \tag{6.3.1}$$

with finite positive ordinate of convergence τ, such that $C(z)$ is an entire function. By virtue of Theorem 6.1 the associated Hermitian series

$$f(z) = \sum_{n=0}^{\infty} f_{2n} h_{2n}(z) \tag{6.3.2}$$

also represents an entire function and its ordinate of convergence equals τ.

Let $g(w)$ be an entire function of order $\frac{1}{2}$ and normal type τ, having only real negative zeros, and form the series

$$C(z) = \sum_{n=0}^{\infty} (-1)^n \frac{\cos (4n + 1)^{\frac{1}{2}} z}{g(4n + 1)}. \tag{6.3.3}$$

For sufficiently large values of r we have

$$\exp [(\tau - \epsilon) r^{\frac{1}{2}}] \leqq g(r) \leqq \exp [(\tau + \epsilon) r^{\frac{1}{2}}].$$

Consequently the ordinate of convergence of the series is τ. On the other hand, form the integral[31]

$$I(z) = \int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} \frac{\cos (4t + 1)^{\frac{1}{2}} z}{g(4t + 1)} \cdot \frac{e^{\pi i t}}{e^{2\pi i t} - 1}\, dt. \tag{6.3.4}$$

It is easy to see that $I(z)$ is uniformly convergent in any bounded domain so that $I(z)$ is an entire function. Familiar methods in the calculus of residues show that $I(z) \equiv C(z)$ within the domain of convergence of the series. Hence

$$f(z) = \sum_{n=0}^{\infty} \frac{h_{2n}(z)}{A_{2n}\, g(4n + 1)} \tag{6.3.5}$$

is a Hermitian series whose ordinate of convergence is τ, $0 < \tau < \infty$, and which represents an entire function.

6.4. **Continuable gap-series.** In order to bring out the limits of possible improvement in the gap theorems proved in §5.5 we shall prove

THEOREM 6.3. *Let*

$$g(w) = \sum_{k=0}^{\infty} a_{2k+1} w^{2k+1} \tag{6.4.1}$$

be a power series, convergent for $|w| < R$, $R > 1$, *having at least one arc of regularity on the circle of convergence. Then*

$$f(z) = \sum_{n=0}^{\infty} \frac{a_{2n+1}}{A_{k_n}} h_{k_n}(z), \qquad k_n = 2n^2 + 2n, \tag{6.4.2}$$

is a Hermitian gap-series having log R *as its ordinate of convergence, and admitting at least one interval of regularity on each line of convergence.*

[31] The path of integration is supposed to lie in the left half plane but to the right of the zeros of $g(4t + 1)$.

Proof. The assumptions on $g(w)$ imply that the series

$$C(z) = \sum_{k=0}^{\infty} a_{2k+1} \cos (2k+1)z \tag{6.4.3}$$

is convergent in the strip $-\log R < y < \log R$ and has at least one interval of regularity on each of the lines of convergence. $C(z)$ being an even function of z, we conclude that these intervals are also intervals of regularity of the associated Hermitian series. We have then merely to notice that (6.4.2) is actually the associated Hermitian series of (6.4.3).

COROLLARY. *The condition $n_k = O(k^2)$ is not sufficient to make the gap-series*

$$\sum_{n=1}^{\infty} f_{n_k} h_{n_k}(z)$$

non-continuable.

6.5. **Expression of a c_n-series in terms of its associated Hermitian series.** We now take up our last problem, that of expressing $C(z)$ in terms of $f(z)$. Here we obtain a much less satisfactory solution than in the converse problem. Our point of departure is again the formula

$$h_n(z) = c_n(z) + \frac{1}{\nu_n} \int_0^z t^2 \sin \nu_n(z-t)\, h_n(t)\, dt. \tag{6.5.1}$$

We multiply both sides by f_n and sum for n. The result can be written

$$f(z) = C(z;f) + \int_0^z t^2 T(z-t, t; f)\, dt, \tag{6.5.2}$$

where

$$T(w, t; f) = \sum_{n=0}^{\infty} \frac{1}{\nu_n} f_n \sin (\nu_n w)\, h_n(t) \tag{6.5.3}$$

and the integral is taken along a straight line.

Putting $z = x + iy$, $t = u + iv$, and recalling that

$$| h_n(t) | \, e^{-\nu_n |v|} A_n^{-1} \to 1 \quad \text{as} \quad n \to \infty,$$

we conclude that the series for $T(z - t, t; f)$ is dominated by a suitable multiple of the series

$$\sum_{n=0}^{\infty} \frac{1}{\nu_n} | f_n | \, A_n e^{\nu_n |y|}.$$

It is consequently convergent for $-\tau < y < \tau$, where τ as usual is the ordinate of convergence of the Hermitian series for $f(z)$. It follows that formula (6.5.2) is valid in $-\tau < y < \tau$.

Suppose now that $f(z)$ can be continued analytically along a finite path from

$z = 0$ to a point outside of the strip of convergence. Under what circumstances can $C(z)$ be continued along the same path? Now

$$\sin\left[(2n+1)^{\frac{1}{2}}(z-t)\right] h_n(t) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!}(2n+1)^{k+\frac{1}{2}}(z-t)^{2k+1}h_n(t)$$

$$= (2n+1)^{\frac{1}{2}} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!}(z-t)^{2k+1}\delta_t^k h_n(t),$$

so that

$$T(z-t,t;f) = \sum_{n=0}^{\infty} f_n \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!}(z-t)^{2k+1}\delta_t^k h_n(t). \tag{6.5.4}$$

This double series is dominated by a suitable multiple of

$$\sum_{n=0}^{\infty} (2n+1)^{-\frac{1}{4}} A_n \,|f_n|\, \sinh\left[(2n+1)^{\frac{1}{2}}\,|z-t|\right] \exp\left[(2n+1)^{\frac{1}{2}}\,|v|\right]$$

which is absolutely convergent for $|z-t| + |v| < \tau$. But t has to describe a path in its plane from $t = 0$ to $t = z$. This, in particular, requires $|z| < \tau$, and if this condition is fulfilled, $|z-t| + |v| < \tau$ along the straight line joining $t = 0$ with $t = z$.

Hence (6.5.4) is absolutely convergent for $|z| < \tau$, $t = \rho z$, $0 \leqq \rho \leqq 1$. We can then rearrange the series and obtain

$$T(z-t,t;f) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!}(z-t)^{2k+1}\delta_t^k f(t). \tag{6.5.5}$$

Symbolically we can write

$$T(z-t,t;f) = \frac{\sin\left[(z-t)\delta_t^{\frac{1}{2}}\right]}{\delta_t^{\frac{1}{2}}} f(t). \tag{6.5.6}$$

Suppose now that z is a point of the principal Mittag-Leffler star of $f(z)$ and let t_0 be a point on the line segment joining z with the origin. Let the radius of holomorphism of $f(t)$ at $t = t_0$ be $R(t_0)$. Then by formula (3.3.12) and footnote 6

$$|\delta_t^k f(t)| \leqq B(\epsilon)(2k)!\,[R(t) - \epsilon]^{-2k},$$

where $B(\epsilon)$ is independent of k and of t but of course depends upon $f(z)$. It follows that

$$|T(z-t,t;f)| \leqq B(\epsilon) \sum_{k=0}^{\infty} \frac{1}{2k+1}\,|z-t|^{2k+1}\,[R(t) - \epsilon]^{-2k}.$$

This series converges for $|z-t| < R(t) - \epsilon$. In other words, z must be nearer to t than t is to the boundary of the domain of holomorphism.

In order that we shall be sure that

$$T(z;f) \equiv \int_0^z t^{\frac{1}{2}} T(z-t,t;f)\,dt = f(z) - C(z;f) \tag{6.5.7}$$

be holomorphic, the above-mentioned condition must be valid all along the path of integration. For $t = 0$ this condition implies in particular that $f(t)$ is holomorphic in a circle with center at the origin containing the point z under consideration. On the other hand, if this special condition is satisfied, then we can use a rectilinear path of integration. With such a path the local condition of convergence is satisfied everywhere along the path and $T(z; f)$ is holomorphic, hence also $C(z; f)$. Thus we have proved

THEOREM 6.4. *If $f(z)$ is holomorphic in the circle $|z| < R$, then its associated functions $C(z; f)$, $S(z; f)$, $E^+(z; f)$, and $E^-(z; f)$ are holomorphic in the same circle.*

Combining this theorem with Theorem 6.1, we get some interesting results which we formulate as

THEOREM 6.5. *If $z = z_0$ is the singular point of $C(z; f)$ which is nearest to the origin, or one out of several such points, then either z_0 or $-z_0$ is a singular point of $f(z)$. Conversely, every singular point of $f(z)$ is a singular point of either $C(z; f)$ or of $C(-z; f)$. In particular, $f(z)$ is an entire function if and only if $C(z; f)$ is also entire.*

It would be interesting to know if all the singular points of $C(z; f)$ give rise to singular points of $f(z)$. There is no direct evidence favoring such a hypothesis, but it seems very plausible. It is possible that a more refined analysis of the properties of the function $T(z - t, t; f)$ would decide this question.

Addendum (September 15, 1939). After the present paper had been received by the editors, there appeared a basic treatise on orthogonal polynomials by G. Szegö (American Mathematical Society Colloquium Publications, vol. XXIII). This treatise has numerous points of contact with Chapters 1 and 2 of the present paper. Szegö uses the method of Chapter 1, which he refers to as that of Liouville-Stekloff, to derive several asymptotic relations. His results on Hermitian polynomials are stated in Theorems 8.22.6 and 8.22.7 (pp. 193–194, comments on pp. 196–198, proof on pp. 212–213) and should be compared with our formulas (1.2.8), (1.2.9) and Theorem 1.6 which give somewhat additional information. On page 246 Szegö states that the domain of convergence of a Hermitian series is a horizontal strip and refers to his formula (8.23.4) on page 197 for a proof.

BIBLIOGRAPHY

1. A. ADAMOFF, *Sur les expansions des polynomes $U_n = e^{\frac{1}{2}ax^2} d^n(e^{-\frac{1}{2}ax^2})/dx^n$ pour les grands valeurs de n* (in Russian), Annales de l'Institut Polytechnique de St. Pétersbourg, vol. 5(1906), pp. 127–143.
2. F. CARLSON AND E. LANDAU, *Neuer Beweis und Verallgemeinerungen des Fabryschen Lückensatzes*, Göttinger Nachrichten, Math.-Phys. Kl., (1921), pp. 184–188.
3. H. CRAMÉR, *Sur une classe de séries de Dirichlet*, Thèse, Upsala, 1917, 51 pp.
4. H. CRAMÉR, *Un théorème sur les séries de Dirichlet et son application*, Arkiv för Mat., Astr. o. Fys., vol. 13(1918), 14 pp.
5. P. DIENES, *Essai sur les singularités des fonctions analytiques*, Journal de Math. pures et appl., (6), vol. 5(1909), pp. 327–413.

6. G. Faber, *Über die Fortsetzbarkeit gewisser Taylorschen Reihen*, Math. Annalen, vol. 57(1903), pp. 369–388.
7. G. Faber, *Über die Nichtfortsetzbarkeit gewisser Potenzreihen*, Münchner Sitzungsberichte, Math.-Phys. Kl., vol. 34(1904), pp. 63–74.
8. E. Fabry, *Sur les points singuliers d'une fonction donnée par son développement en série et l'impossibilité du prolongement analytique dans des cas très généraux*, Ann. Ecole Norm. Sup., (3), vol. 13(1896), pp. 367–399.
9. E. Fabry, *Sur les séries de Taylor qui ont une infinité de points singuliers*, Acta Math., vol. 22(1899), pp. 65–87.
10. P. Fatou, *Séries trigonométriques et séries de Taylor*, Acta Math., vol. 30(1906), pp. 335–400.
11. E. Hille, *On the zeros of the functions of the parabolic cylinder*, Arkiv för Mat., Astr. o. Fys., vol. 18(1924), 56 pp.
12. E. Hille, *A class of reciprocal functions*, Annals of Math., (2), vol. 27(1926), pp. 427–464.
13. A. Hurwitz and G. Pólya, *Zwei Beweise eines von Herrn Fatou vermuteten Satzes*, Acta Math., vol. 40(1915), pp. 179–183.
14. L. Leau, *Recherche des singularités d'une fonction définie par un développement de Taylor*, Journal de Math., (5), vol. 5(1899), pp. 365–425.
15. H. Muggli, *Differentialgleichungen unendlich hoher Ordnung mit konstanten Koeffizienten*, Commentarii Math. Helvetici, vol. 11(1938), pp. 151–179.
16. A. Ostrowski, *Über vollständige Gebiete gleichmässiger Konvergenz von Folgen analytischer Funktionen*, Abhandlungen Math. Sem. Hamburg, vol. 1(1921–22), pp. 327–350.
17. A. Ostrowski, *Über einen Satz von Leau und die analytische Fortsetzung einiger Klassen Taylorscher und Dirichletscher Reihen*, Math. Annalen, vol. 108(1933), pp. 718–756.
18. M. Plancherel and W. Rotach, *Sur les valeurs asymptotiques des polynomes d'Hermite*, Commentarii Math. Helvetici, vol. 1(1929), pp. 227–254.
19. G. Pólya, *Untersuchungen über Lücken und Singularitäten von Potenzreihen*, Math. Zeitschrift, vol. 29(1929), pp. 549–640.
20. A. Pringsheim, *Über Funktionen, welche in gewissen Punkten endliche Differentialquotienten jeder endlicher Ordnung, aber keine Taylorsche Reihenentwicklung besitzen*, Math. Annalen, vol. 44(1894), pp. 41–56.
21. A. Pringsheim, *Über einige funktionentheoretische Anwendungen der Eulerschen Reihen-Transformation*, Münchner Sitzungsberichte, Math.-Phys. Kl., 1912, pp. 11–91.
22. N. Schwid, *The asymptotic forms of the Hermite and Weber functions*, Transactions Amer. Math. Soc., vol. 37(1935), pp. 339–362.
23. O. Szász, *Über Singularitäten von Potenzreihen und Dirichletschen Reihen am Rande des Konvergenzbereiches*, Math. Annalen, vol. 85(1922), pp. 99–110.
24. G. Szegö, *Beiträge zur Theorie der Laguerreschen Polynome*, I, *Entwicklungssätze*, Math. Zeitschrift, vol. 25(1926), pp. 87–115.
25. G. Valiron, *Lectures on the general theory of integral functions*, Toulouse, 1923, xi + 208 pp.
26. G. Vitali and G. Sansone, *Moderna teoria delle funzioni di variabile reale*, Parte seconda, Bologna, 1935, vi + 310 pp.
27. G. Vivanti, *Sulle serie di potenze*, Revista di Matematica, vol. 3(1893), pp. 111–114.
28. O. Volk, *Über die Entwicklung von Funktionen einer komplexen Veränderlichen nach Funktionen, die einer linearen Differentialgleichung zweiter Ordnung genügen*, Math. Annalen, vol. 86(1922), pp. 296–316.
29. O. Volk, *Über die Entwicklung komplexer Funktionen nach den Hermiteschen und Laguerreschen Funktionen*, Mémoires Fac. Sci. Université de Lithuanie, Kaunas, 1923, pp. 3–33.

30. G. N. Watson, *The harmonic functions associated with the parabolic cylinder*, Proc. London Math. Soc., (2), vol. 8(1910), pp. 393–421.
31. G. N. Watson, *The harmonic functions associated with the parabolic cylinder*, Proc. London Math. Soc., (2), vol. 17(1918), pp. 116–148.
32. E. T. Whittaker, *On the functions associated with the parabolic cylinder in harmonic analysis*, Proc. London Math. Soc., vol. 35(1903), pp. 417–427.
33. S. Wigert, *Quelques théorèmes sur les fonctions entières*, Öfversigt K. Vetenskaps-Akademiens Förhandlingar, Stockholm, 1902, pp. 207–214.

Yale University.

A UNIQUENESS THEOREM FOR ANALYTIC ALMOST-PERIODIC FUNCTIONS

By S. Bochner

We start from the following theorem which can be proved along familiar lines.

(i) If the functions $F_n(z)$ $(n = 1, 2, \cdots)$ are each analytic in the annulus $a < r < b$ and (uniformly) continuous in the closure $a \leqq r \leqq b$,

(ii) if there exists a constant A such that for $a \leqq r \leqq b$ and $n = 1, 2, \cdots$

$$\frac{1}{2\pi}\int_0^{2\pi} |F_n(re^{i\theta})|\, d\theta \leqq A,$$

and

(iii) if there exists an interval $\theta_0 < \theta < \theta_1$ for which

$$\lim_{n\to\infty} \frac{1}{2\pi}\int_{\theta_0}^{\theta_1} |F_n(ae^{i\theta})|\, d\theta = 0$$

or, what is the same, if there exists a non-negative continuous periodic function $\varphi(\theta)$, which does not vanish identically, for which

$$\lim_{n\to\infty} \frac{1}{2\pi}\int_0^{2\pi} |F_n(ae^{i\theta})|\, \varphi(\theta)\, d\theta = 0,$$

then the sequence of functions $F_n(z)$ converges towards 0 everywhere in the annulus.

Putting $z = e^s$, $s = \sigma + it$, we will generalize this theorem to analytic almost-periodic functions in strips.[1] The periodic function $\varphi(\theta)$ will be replaced by an (uniformly continuous) almost-periodic function of the real variable t. The mean value

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \chi(t)\, dt$$

of an almost-periodic function $\chi(t)$ will be denoted by

$$M_t\chi(t).$$

Theorem. (i) *If each function* $f_n(s)$ $(n = 1, 2, \cdots)$ *is analytic in the strip*

(1) $$\alpha < \sigma < \beta$$

and uniformly continuous and almost periodic in the closed strip

(2) $$\alpha \leqq \sigma \leqq \beta,$$

Received September 13, 1939.

[1] See A. S. Besicovitch, *Almost Periodic Functions*, Cambridge, 1932.

(ii) *if there exists a constant A such that for $\alpha \leqq \sigma \leqq \beta$ and $n = 1, 2, \cdots$*

$$M_t \,|\, f_n(\sigma + it) \,| \leqq A, \tag{3}$$

and

(iii) *if there exists a non-negative almost-periodic function $\varphi(t)$, which does not vanish identically, for which*

$$\lim_{n\to\infty} M_t(|\, f_n(\alpha + it) \,|\, \varphi(t)) = 0, \tag{4}$$

then the relation

$$\lim_{n\to\infty} M_t \,|\, f_n(\sigma + it) \,| = 0 \tag{5}$$

holds for every σ in the open interval (1).

As a matter of fact, the limit relation (5) holds uniformly in every interior interval $\alpha_1 \leqq \sigma \leqq \beta_1$, $\alpha_1 > \alpha$, $\beta_1 < \beta$ since, by a general theorem,[2] the logarithm of the function $M_t \,|\, f_n(\sigma + it) \,|$ is a convex function of σ.

The proof of the theorem will require several steps.

Definition. A point set E of the real axis will be called an (l, δ)-set, $l > \delta > 0$, if the intersection of E with any interval of length l contains an interval of length δ.

The function $\varphi(t)$ of assumption (iii) is (uniformly continuous and) almost periodic. Therefore, if

$$\eta = \tfrac{1}{4} \sup_t \varphi(t),$$

there exists an $h > 0$ such that

$$\varphi(t - \tau) \leqq \varphi(t) + \eta \tag{6}$$

for $|\, \tau \,| \leqq h$ and all t; and there exist an $l_0 > 0$ and a translation number $p = p(\eta)$ in each interval of length l_0 such that

$$\varphi(t - \tau - p) \leqq \varphi(t - \tau) + \eta. \tag{7}$$

From (6) and (7) we conclude that there exists an (l, δ)-set E, with $l = 2l_0$ and $\delta = \min\,(h, l)$, such that

$$\varphi(t - \tau) \leqq \varphi(t) + 2\eta$$

for τ in E and all t. It is now easy to see that the function

$$\psi(t) = \max\,[\varphi(t) - 2\eta, 0]$$

will satisfy the following lemma:

Lemma 1. *Corresponding to the function $\varphi(t)$ of assumption* (iii) *there exist an (l, δ)-set E and another function $\psi(t)$ which is also almost periodic, non-negative and not identically* 0 *such that the relation*

[2] Hardy, Ingham and Pólya, *Theorems concerning mean values of analytic functions*, Proceedings of the Royal Society A, vol. 113(1926), pp. 542-569.

$$\psi(t - \tau) \leqq \varphi(t) \tag{8}$$

holds for τ in E and all t.

We shall next derive a lemma on analytic functions which is based on the following theorem. If the boundary of an open domain D in the s-plane contains a simple arc, if a sequence of functions $h_n(s)$ in D are uniformly bounded and each uniformly continuous in D, and if the boundary values of $h_n(s)$ on the arc are uniformly convergent to 0 as $n \to \infty$, then the sequence is uniformly convergent to 0 in every closed part of D.[3]

LEMMA 2. *If (1) is a given strip in the s-plane and $\sigma = \gamma$ is a fixed interior line, if E is a fixed (l, δ)-set on the boundary line $\sigma = \alpha$, $-\infty < t < \infty$, and if B is a fixed positive constant, then there exists an error function $\rho(\epsilon)$,*

$$\lim_{\epsilon \to 0} \rho(\epsilon) = 0,$$

having the following property:

If $g(s)$ is analytic in (1) and uniformly continuous in (2), if $|g(s)| \leqq B$, and if

$$|g(\alpha + it)| \leqq \epsilon \qquad \textit{for } t \textit{ in } E, \tag{9}$$

then

$$|g(\gamma + it)| \leqq \rho(\epsilon) \qquad \textit{for } -\infty < t < \infty.$$

If the lemma were false there would exist a sequence of functions $g_n(s)$ which are analytic in (1) and uniformly continuous in (2), a sequence of positive numbers ϵ_n, $\epsilon_n \to 0$, a sequence of real numbers t_n and a number $\rho > 0$ such that $|g_n(s)| \leqq B$,

$$|g_n(\alpha + it)| \leqq \epsilon_n \qquad \text{for } t \text{ in } E \tag{10}$$

and

$$|g_n(\gamma + it_n)| \geqq \rho. \tag{11}$$

For every n there exists an interval $p_n \leqq t \leqq p_n + \delta$ belonging to E such that $p_n \leqq t_n \leqq p_n + l$. The functions $h_n(s) = g_n(s - p_n)$ are again uniformly bounded in (1) and each uniformly continuous in (2); and, on account of (10), the sequence $h_n(\alpha + it)$ is uniformly convergent to 0 on the arc $0 \leqq t \leqq \delta$. However, on account of (11) the sequence $h_n(\gamma + it)$ is not uniformly convergent to 0 on the closed set $|t| \leqq l$ and we thus obtain a contradiction to the above-mentioned theorem.

Finally we shall need another simple lemma on almost-periodic functions:

LEMMA 3. *If $F(t)$ is a fixed almost-periodic function and $\chi(t)$ is a variable almost-periodic function for which*

$$|\chi(t)| \leqq 1, \tag{12}$$

then

$$\sup |M_t F(t)\chi(t)| = M_t^* |F(t)|.$$

[3] Compare R. Nevanlinna, *Eindeutige analytische Funktionen*, 1936, Chapter III.

If $F(t)$ is bounded away from 0, the lemma is quite trivial, the sup being attained for

$$\chi(t) = \operatorname{sign} F(t). \tag{13}$$

In the general case the function (13) is a bounded Stepanoff function and the value of the sup for variable bounded Stepanoff functions is the same as for variable almost-periodic functions.

We are now ready for the proof of our theorem. We take a fixed function $\psi(t)$ satisfying Lemma 1 and an arbitrary almost-periodic function satisfying (12), and we transform the functions $f_n(s)$ of the theorem into the new functions

$$g_n(s) = M_\tau f_n(\sigma + it + i\tau)\psi(\tau)\chi(\tau).$$

It is easy to verify that they are again analytic in (1) and almost periodic and uniformly continuous in (2). Moreover, putting

$$B = A \sup_t \varphi(t),$$

$$\epsilon_n = M_t \,|\, f_n(\alpha + it) \,|\, \varphi(t),$$

we obtain, for all (σ, t),

$$|\, g_n(\sigma + it) \,| \leqq M_\tau \,|f_n(\sigma + it + i\tau)\,|\cdot \sup_\tau \chi(\tau) \leqq B$$

and, for all t in E,

$$\begin{aligned} |\, g_n(\alpha + it) \,| &\leqq M_\tau \,|\, f_n(\alpha + it + i\tau) \,|\, \psi(\tau) \\ &= M_\tau \,|\, f_n(\alpha + i\tau) \,|\, \psi(\tau - t) \leqq M_\tau \,|\, f_n(\alpha + i\tau) \,|\, \varphi(\tau) = \epsilon_n . \end{aligned}$$

Thus all assumptions of Lemma 2 are fulfilled and we conclude that the inequality

$$g_n(\gamma + it) = |\, M_\tau f_n(\gamma + it + i\tau)\psi(\tau)\chi(\tau) \,| \leqq \rho(\epsilon_n)$$

holds for all t. By Lemma 3 we may eliminate the variable function $\chi(t)$, and we obtain the estimate

$$M_\tau \,|\, f_n(\gamma + it + i\tau) \,|\, \psi(\tau) \leqq \rho(\epsilon_n) \qquad (-\infty < t < \infty).$$

Hence we have

$$\begin{aligned} \rho(\epsilon_n) &\geqq M_t(M_\tau \,|\, f_n(\gamma + it + i\tau) \,|\, \psi(\tau)) \\ &= M_\tau(M_t \,|\, f_n(\gamma + it + i\tau) \,|\, \psi(\tau)) \\ &= M_t \,|\, f_n(\gamma + it) \,|\cdot M_\tau \psi(\tau). \end{aligned}$$

Since $\rho(\epsilon_n) \to 0$ by (4), we finally obtain relation (5) for every $\sigma = \gamma$ in (2).

PRINCETON UNIVERSITY.

SOME SUMS INVOLVING POLYNOMIALS IN A GALOIS FIELD

By L. Carlitz

1. Let $GF(p^n)$ denote a Galois (finite) field of order p^n. Let

$$M = M(x) = c_0x^m + c_1x^{m-1} + \cdots + c_m$$

denote a polynomial in the indeterminate x with coefficients in the $GF(p^n)$. For $c_0 = 1$, M is said to be primary; provided $c_0 \neq 0$, we write $\deg M = m$. In this note we evaluate certain sums extended over the set of primary polynomials of a given degree. Some of the formulas are new, others were proved earlier but are derived here in a more direct manner. In particular we shall prove the identities

$$\sum_{\deg M=m} M^{p^{n(m+k)}-1} = (-1)^m \frac{F_{m+k}}{L_m F_k^{p^{nm}}}, \tag{1.1}$$

and

$$\sum^{m}_{\deg M=m} \frac{1}{M^{p^{nk}-1}} = \frac{L_{k+m-1}}{L_{k-1} L_m^{p^{nk}}}, \tag{1.2}$$

where

$$\begin{aligned} F_k &= (x^{p^{nk}} - x)(x^{p^{nk}} - x^{p^n}) \cdots (x^{p^{nk}} - x^{p^{n(k-1)}}), \\ L_k &= (x^{p^{nk}} - x)(x^{p^{n(k-1)}} - x) \cdots (x^{p^n} - x), \\ F_0 &= L_0 = 1. \end{aligned} \tag{1.3}$$

2. We shall require a number of known formulas.[1] Put

$$\psi_m(t) = \sum_{j=0}^{m} (-1)^{m-j} \begin{bmatrix} m \\ j \end{bmatrix} t^{p^{nj}}, \tag{2.1}$$

where

$$\begin{bmatrix} m \\ j \end{bmatrix} = \frac{F_m}{F_j L_{m-j}^{p^{nj}}}, \qquad \begin{bmatrix} m \\ 0 \end{bmatrix} = \frac{F_m}{L_m}, \qquad \begin{bmatrix} m \\ m \end{bmatrix} = 1.$$

Then $\psi_m(G) = 0$ for all G of degree less than m, while $\psi_m(M) = F_m$ for M primary of degree equal to m; indeed we have the factorizations

$$\psi_m(t) = \prod_{\deg G < m} (t + G) \tag{2.2}$$

Received October 12, 1939.

[1] *On certain functions connected with polynomials in a Galois field*, this Journal, vol. 1(1935), pp. 137–168, p. 139. This paper will be cited as DJ.

taken over all polynomials (including 0) of degree less than m, and

$$\psi_m(t) + F_m = \prod_{\deg M = m} (t + M), \tag{2.3}$$

taken over all primary polynomials of degree m.

From (2.3) follows

$$(-1)^m \frac{F_m}{L_m} \frac{1}{\psi_m(t) + F_m} = \sum_{\deg M = m} \frac{1}{t + M}, \tag{2.4}$$

and from this we get (putting $t = 0$)

$$\sum_{\deg M = m} \frac{1}{M} = \frac{(-1)^m}{L_m}. \tag{2.5}$$

3. In order to prove (1.1) we require the closely related formula

$$\sum_{\deg M = m} M^{p^{nk}-1} = 0 \qquad \text{for } k < m. \tag{3.1}$$

This may be proved by expanding both sides of (2.4) in descending powers of t, but for our purpose a different sort of proof seems preferable. Consider therefore the sum

$$\sum_{\deg M = m} \frac{\psi_k(M)}{M} \qquad (k < m), \tag{3.2}$$

taken over primary M of degree m. We write $M = x^k A + B$, where A and B run over the primary polynomials of degree $m - k$ and k, respectively. Then the sum in (3.2)

$$= \sum_{\deg A = m-k} \sum_{\deg B = k} \frac{\psi_k(x^k A + B)}{x^k A + B} = \sum_A \sum_B \frac{\psi_k(x^k A) + F_k}{x^k A + B}$$

$$= \sum_A \{\psi_k(x^k A) + F_k\} \cdot (-1)^k \frac{F_m}{L_m} \frac{1}{\psi_k(x^k A + B)},$$

by (2.4). Clearly the last sum vanishes; thus

$$\sum_{\deg M = m} \frac{\psi_k(M)}{M} = 0 \qquad \text{for } k < m. \tag{3.3}$$

But by (2.1), this identity may be put in the form

$$\sum_{j=0}^{k} (-1)^{k-j} \begin{bmatrix} k \\ j \end{bmatrix} \sum_M M^{p^{nj}-1} = 0,$$

and from this (3.1) follows at once.

We remark that (3.3) can easily be extended to cover the case $k \geq m$. For $k > m$, $\psi_k(M) = 0$ and thus (3.3) holds. For $k = m$, $\psi_k(M) = F_m$, and we have

$$\sum_{\deg M = m} \frac{\psi_m(M)}{M} = \sum_M \frac{F_m}{M} = (-1)^m \frac{F_m}{L_m},$$

by (2.5). Hence we may replace (3.3) by

$$\sum_{\deg M=m} \frac{\psi_k(M)}{M} = \begin{cases} 0 & \text{for } k \neq m, \\ (-1)^m \dfrac{F_m}{L_m} & \text{for } k = m. \end{cases} \tag{3.4}$$

4. We shall also require the identity[2]

$$t^{p^{nk}} = \sum_{j=0}^{k} \frac{F_k}{F_j F_{k-j}^{p^{nj}}} \psi_j(t). \tag{4.1}$$

In this identity put $t = M$, and we get

$$\sum_{\deg M=m} M^{p^{nk}-1} = \sum_{j=0}^{k} \frac{F_k}{F_j F_{k-j}^{p^{nj}}} \sum_{M} \frac{\psi_j(M)}{M}.$$

We now apply (3.4) and we have at once

$$\sum_{\deg M=m} M^{p^{nk}-1} = \begin{cases} 0 & \text{for } k < m, \\ (-1)^m \dfrac{F_k}{L_m F_{k-m}^{p^{nm}}} & \text{for } k \geqq m, \end{cases} \tag{4.2}$$

so that we have proved (1.1).

The identity (4.1) can be extended considerably. Consider, for example, the closely related identity[2]

$$\psi_k(tu) = \sum_{j=0}^{k} \frac{F_k}{F_j F_{k-j}^{p^{nj}}} \psi_j(t) \psi_{k-j}^{p^{nj}}(u). \tag{4.3}$$

Again put $t = M$, where $\deg M = m \leqq k$, and we get

$$\sum_{\deg M=m} \frac{\psi_k(Mu)}{M} = \sum_{j=0}^{k} \frac{F_k}{F_j F_{k-j}^{p^{nj}}} \psi_{k-j}^{p^{nj}}(u) \sum_{M} \frac{\psi_j(M)}{M}.$$

Now apply (3.4), and as before we have

$$\sum_{\deg M=m} \frac{\psi_k(Mu)}{M} = (-1)^m \frac{F_k}{L_m F_{k-m}^{p^{nm}}} \psi_{k-m}^{p^{nm}}(u) \tag{4.4}$$

for $k \geqq m$. For $u = 1$ this reduces to (3.4).

The most general formula like (4.2) or (4.4) follows from the identity[2]

$$g(tu) = \sum_{j=0}^{k} \frac{1}{F_j} \Delta^j g(u) \psi_j(t), \tag{4.5}$$

where $g(t)$ is a "linear" polynomial in t:

$$g(t) = A_0 t + A_1 t^{p^n} + \cdots + A_k t^{p^{nk}},$$

$$\Delta g(t) = g(xt) - x g(t),$$

[2] DJ, p. 144.

and generally

$$\Delta^{j+1}g(t) = \Delta^j g(xt) - x^{p^{nj}}\Delta^j(t).$$

From (4.5) follows

$$\sum_{\deg M=m} \frac{g(Mu)}{M} = \frac{(-1)^m}{L_m}\Delta^m g(u), \tag{4.6}$$

and this relation includes both (4.2) and (4.4). We may also look on (4.6) as giving a new expression for $\Delta^m g(u)$.

5. We shall now develop some identities of a different sort. We begin with[3]

$$\psi_{m+1}(t) + F_m^{p^n-1}\cdot\psi_m(t) = \psi_m^{p^n}(t), \tag{5.1}$$

which relation follows directly from (2.1). From (5.1) we get—writing ψ_m for $\psi_m(t)$—

$$\psi_m^{p^{2n}} = \psi_{m+1}^{p^n} + F_m^{p^{2n}-p^n}\psi_m^{p^n} = \psi_{m+2} + (F_{m+1}^{p^n-1} + F_m^{p^{2n}-p^n})\psi_{m+1} + F_m^{p^{2n}-1}\psi_m .$$

If we proceed in this way, it is clear that we get a formula of the following type:

$$\psi_m^{p^{n(k-1)}}(t) = \psi_{m+k-1}(t) + A_1\psi_{m+k-2}(t) + \cdots + A_{k-1}\psi_m(t). \tag{5.2}$$

It is easy to define the A_j in (5.2) recursively. However, we need only A_{k-1}. In (5.2) put $t = x^m$; since $\psi_{m+j}(x^m) = 0$ for $j > 0$, we have at once

$$A_{k-1} = F_m^{p^{n(k-1)}-1}. \tag{5.3}$$

6. Formula (5.2) may be looked on as a linear expression in $\psi_m, \psi_{m+1}, \cdots, \psi_{m+k-1}$ in which the terms in

$$t, t^{p^n}, \cdots, t^{p^{n(k-2)}}$$

are all missing. We shall require a linear expression in $\psi_m, \cdots, \psi_{m+k-1}$ in which the terms in

$$t^{p^n}, t^{2p^n}, \cdots, t^{p^{n(k-1)}}$$

are all missing. It is fairly clear that this exists, and may indeed be derived from (5.2). For replace m by $m - 1$ in the latter formula, raise both sides to the p^n-th power, and use

$$\Delta\psi_k(t) = [k]\psi_{k-1}^{p^n}(t);$$

we get[4]

$$\frac{\psi_{m+k-1}}{[m+k-1]} + A_1^{p^n}\frac{\psi_{m+k-2}}{[m+k-2]} + \cdots + \frac{F_{m-1}^{p^{nk}}}{F_m}\psi_m = Bt + \sum_{j=0}^{m-1} B_j t^{p^{k+j}}, \tag{6.1}$$

[3] DJ, p. 141.

[4] Note that if $\Delta f(t) = \Delta g(t)$, where $f(t)$ and $g(t)$ are linear polynomials, then $f(t) = g(t) + Ct$; that is, $f(t)$ and $g(t)$ can differ only by a term in t.

where for brevity we put $[k] = x^{p^{nk}} - x$. From (2.1) it follows that

$$B_j = (-1)^{m-1-j} \begin{bmatrix} m-1 \\ j \end{bmatrix}^{p^{nk}} \frac{1}{[k+j]}, \tag{6.2}$$

but this will not be needed for the application. It remains to determine B.

In (6.1) put $t = x^i$ $(i = 0, 1, \cdots, m)$. We get the set of equations

$$Bx^i + \sum_{j=0}^{m-1} B_j x^{ip^{k+j}} = \begin{cases} 0 & \text{for } i = 0, 1, \cdots, m-1, \\ F_{m-1}^{p^{nk}} & \text{for } i = m. \end{cases} \tag{6.3}$$

The determinant of the left member of (6.3) is

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x & x^{p^{nk}} & \cdots & x^{p^{n(k+m-1)}} \\ \cdots & \cdots & \cdots & \cdots \\ x^m & x^{mp^{nk}} & \cdots & x^{mp^{n(k+m-1)}} \end{vmatrix}.$$

This is a Vandermonde determinant and therefore[5]

$$\begin{aligned} &= \prod_{j=0}^{m-1} (x^{p^{n(k+j)}} - x) \cdot \prod_{0 \le i < j < m} (x^{p^{n(k+j)}} - x^{p^{n(k+i)}}) \\ &= \frac{L_{k+m-1}}{L_{k-1}} (F_1 F_2 \cdots F_{m-1})^{p^{nk}}, \end{aligned} \tag{6.4}$$

by (1.3). In the next place the numerator for B is evidently

$$(-1)^m F_{m-1}^{p^{nk}} \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x^{p^{nk}} & x^{p^{n(k+1)}} & \cdots & x^{p^{n(k+m-1)}} \\ \cdots & \cdots & \cdots & \cdots \\ x^{(m-1)p^{nk}} & x^{(m-1)p^{n(k+1)}} & \cdots & x^{(m-1)p^{n(k+m-1)}} \end{vmatrix} \tag{6.5}$$

$$= (-1)^m F_{m-1}^{p^{nk}} (F_1 F_2 \cdots F_{m-1})^{p^{nk}}.$$

Comparing (6.4) and (6.5) we get

$$B = (-1)^m F_{m-1}^{p^{nk}} \frac{L_{k-1}}{L_{k+m-1}}. \tag{6.6}$$

We remark that in the same way it is easy to solve the system (6.3) for B_j; this leads again to (6.2).

7. Returning to (6.1) we put $t = M$, and now divide by $M^{p^{nk}}$, where M is of degree m. Then all terms but one on the left vanish and (6.1) implies

$$F_{m-1}^{p^{nk}} \sum_{\deg M = m} \frac{1}{M^{p^{nk}}} = B \sum_M \frac{1}{M^{p^{nk}-1}} + \sum_{j=0}^{m-1} B_j \{\sum_M M^{p^{nj}-1}\}^{p^{nk}}. \tag{7.1}$$

[5] Compare DJ, p. 140.

But by (3.1)

$$\sum_M M^{p^{nj}-1} = 0 \qquad \text{for } 0 \leqq j \leqq m-1;$$

also by (2.5)

$$\sum_M \frac{1}{M^{p^{nk}}} = \frac{(-1)^m}{L_m^{p^{nk}}};$$

therefore (7.1) becomes

$$B \sum_M \frac{1}{M^{p^{nk}-1}} = (-1)^m \left(\frac{F_{m-1}}{L_m}\right)^{p^{nk}}.$$

If now we compare with (6.6), we have at once

$$\sum_{\deg M = m} \frac{1}{M^{p^{nk}-1}} = \frac{L_{k+m-1}}{L_{k-1} L_m^{p^{nk}}}, \tag{7.2}$$

so that we have proved (1.2).

8. Consider now the identity ($m > 0$)

$$\frac{L_{k+m}}{L_k L_m^{p^{nk}}} - \frac{L_{k+m-1}}{L_k L_{m-1}^{p^{nk}}} = \frac{L_{k+m-1}}{L_{k-1} L_m^{p^{nk}}}, \tag{8.1}$$

which follows immediately from the definition (1.3). (For $m = 0$ the second term on the left of (8.1) may be ignored and the identity still holds.) If we sum with respect to m, (8.1) implies

$$\sum_{j=0}^{m} \frac{L_{k+j-1}}{L_{k-1} L_j^{p^{nk}}} = \frac{L_{k+m}}{L_k L_m^{p^{nk}}}.$$

Thus comparison with (7.2) leads to[6]

$$\sum_{\deg M \leqq m} \frac{1}{M^{p^{nk}-1}} = \frac{L_{k+m}}{L_k L_m^{p^{nk}}}, \tag{8.2}$$

the sum on the left now extending over all primary polynomials of degree $\leqq m$.

Finally if we put

$$\frac{L_{k+m}}{L_m^{p^{nk}}} = \frac{L_{k+m}}{(x^{p^n} - x)^{p^{n(k+m)}/(p^n-1)}} \left\{\frac{(x^{p^n} - x)^{p^{nm}/(p^n-1)}}{L_m}\right\}^{p^{nk}},$$

and let m become infinite, (8.2) becomes[6]

$$\sum_M \frac{1}{M^{p^{nk}-1}} = \frac{\xi^{p^{nk}-1}}{L_k}, \tag{8.3}$$

where

$$\xi = \lim_{m\to\infty} \frac{(x^{p^n} - x)^{p^{nm}/(p^n-1)}}{L_m},$$

and the sum on the left of (8.3) is taken over *all* primary polynomials.

[6] DJ, p. 161.

9. It is of some interest to remark that for $m = 1$ the formulas (1.1) and (1.2) may be proved very quickly. Indeed, consider the identity

$$(x + c)^{p^{nk}} - (x + c) = [k], \tag{9.1}$$

where c is in $GF(p^n)$. If we divide both sides by $x + c$ and sum over c, we have at once

$$\sum_c (x + c)^{p^{nk}-1} = -\frac{[k]}{[1]} = -\frac{F_k}{L_1 F_{k-1}^{p^n}},$$

and this agrees with (1.1). On the other hand, if we divide both sides of (9.1) by $(x + c)^{p^{nk}}$, we get

$$\sum_c \frac{1}{(x + c)^{p^{nk}-1}} = \frac{[k]}{[1]^{p^{nk}}} = \frac{L_k}{L_{k-1} L_1^{p^{nk}}},$$

and this checks with (1.2).

For $m > 1$ this method will not furnish the final formulas; instead we get certain identities. The generalized version of (9.1) is

$$\psi'_m(M) = F'_m, \tag{9.2}$$

where $\psi'_m(t)$ and F'_m are defined relative to $GF(p^{nk})$. That is, in place of (1.3) we put

$$F'_m = [m]'[m - 1]'^{p^{nk}} \cdots [1]'^{p^{nk(m-1)}},$$

$$L'_m = [m]'[m - 1]' \cdots [1]',$$

where

$$[m]' = x^{p^{mnk}} - x.$$

It is now clear how $\begin{bmatrix} m \\ j \end{bmatrix}'$ and $\psi'_m(t)$ may be defined. Thus (9.2) leads to such identities as

$$\sum_{j=1}^{m} (-1)^{m-j} \begin{bmatrix} m \\ j \end{bmatrix}' \sum_{\deg M = m} M^{p^{nkj}-1} = (-1)^m \frac{F'_m}{L_m}.$$

DUKE UNIVERSITY.

INTEGERS WHICH ARE NOT REPRESENTED BY CERTAIN TERNARY QUADRATIC FORMS

By R. D. James

1. **Introduction.** It is well known that all integers except those of a specified form are represented by certain ternary quadratic forms. Thus, an integer N is represented by the form

$$x^2 + y^2 + z^2$$

if and only if N is not of the form $4^k(8l + 7)$. In this paper we shall show that all sufficiently large integers just fail to be represented by a ternary quadratic form. More precisely, we prove the following[1]

Theorem. *Let any negative integer d be written as $-2^b S^2 h$, where S and h are odd, and h is quadratfrei. Then, every sufficiently large integer N can be represented in the form*

$$N = u^2 + epH^2M, \tag{1.1}$$

where u is a positive integer, p is a prime such that $(d \mid p) = -1$, every prime dividing H also divides d, every prime q dividing M satisfies $(d \mid q) = 1$, and e is 1 except in the three following cases:

$$b \text{ odd}, \quad h \mid N, \quad h \equiv 3 \pmod 4, \quad N \equiv 6 \pmod 8, \tag{1.21}$$

$$b \text{ even}, \quad h \mid N, \quad h \equiv 1 \pmod 4, \quad N \equiv 2 \pmod 4, \tag{1.22}$$

$$b \text{ odd}, \quad h \mid N, \quad h \equiv 1 \pmod 4, \quad N \equiv 2 \pmod 8, \tag{1.23}$$

in all of which $e = 2$.

It is the presence of the single prime p in (1.1) which prevents N from being represented by a ternary quadratic form. For, if $(d \mid p)$ were equal to $+1$, the product pH^2M would be represented by some binary quadratic form of discriminant d.[2] For example, if $d = -3$ the p in (1.1) would be congruent to 2 (mod 3), $H = 3^k$, $k \geqq 0$, and every prime q dividing M would be congruent to 1 (mod 3). The binary quadratic form in question is

$$v^2 + vw + w^2.$$

Received July 14, 1939; in revised form, October 13, 1939.

[1] The author is indebted to the referee for helpful suggestions which led to the proof of the theorem in its present general form.

[2] L. E. Dickson, *Introduction to the Theory of Numbers*, Chapter V.

If the single prime p were congruent to 1 (mod 3), we should have

$$N = u^2 + v^2 + vw + w^2.$$

This ternary quadratic form represents all integers not of the form $9^k(9l + 6)$.

It is easy to see that the conditions (1.21), (1.22), (1.23) are necessary in some cases. Thus when $d = -4$ we have $S = h = 1$. The exceptional case here is $N \equiv 2 \pmod 4$. No such integer can be represented in the form (1.1) with $e = 1$, since $u^2 \equiv 0$ or $1 \pmod 4$, $pH^2M \equiv 0$ or $3 \pmod 4$, and

$$u^2 + pH^2M \not\equiv 2 \pmod 4.$$

The same situation arises if $d = -8$.

The method of proof is an adaptation of the sieve method of Viggo Brun with improvements due to Rademacher[3] and Estermann.[4]

2. **Results from prime number theory.** The following lemmas will be required in the proof of the main theorem.

LEMMA 1. *Let d and D be any two integers for which the Kronecker symbols*[5] *$(d \mid m)$, $(D \mid m)$ are defined. Then if $dD < 0$, we have*

$$\lim_{\xi \to \infty} \sum_{\substack{p \leq \xi \\ (d \mid p) = -1}} \frac{(D \mid p)}{p} = B,$$

where B is a constant depending on d and D.

Proof. The product $(d \mid m)(D \mid m)$ is a character[5] modulo $| dD |$. Also, for $m = | dD | - 1$ we have[5]

$$(d \mid m)(D \mid m) = -1.$$

This means that $(d \mid m)(D \mid m)$ is not the principal character modulo $| dD |$ and hence[6]

$$\sum_{p \leq \xi} \frac{(d \mid p)(D \mid p)}{p} = B_1 + O\left(\frac{1}{\log \xi}\right).$$

This may be written as

$$\sum_{\substack{p \leq \xi \\ (d \mid p) = 1}} \frac{(D \mid p)}{p} - \sum_{\substack{p \leq \xi \\ (d \mid p) = -1}} \frac{(D \mid p)}{p} = B_1 + O\left(\frac{1}{\log \xi}\right). \tag{2.11}$$

[3] Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 3(1924), pp. 12–30.

[4] Journal für die reine und angewandte Mathematik, vol. 168(1932), pp. 106–116.

[5] E. Landau, *Vorlesungen über Zahlentheorie*, vol. 1, pp. 51–56 and pp. 83–87.

[6] E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen*, vol. 1, §109.

Next, $(D \mid m)$ is a character modulo $\mid D \mid$ and is not the principal character. Hence

$$\sum_{p \leq \xi} \frac{(D \mid p)}{p} = B_2 + O\left(\frac{1}{\log \xi}\right)$$

or

$$\sum_{\substack{p \leq \xi \\ (d \mid p) = 1}} \frac{(D \mid p)}{p} + \sum_{\substack{p \leq \xi \\ (d \mid p) = -1}} \frac{(D \mid p)}{p}$$

$$(2.12) \qquad = B_2 - \sum_{p \mid d} \frac{(D \mid p)}{p} + \sum_{\substack{p > \xi \\ p \mid d}} \frac{(D \mid p)}{p} + O\left(\frac{1}{\log \xi}\right).$$

Subtracting (2.11) from (2.12), we obtain

$$2 \sum_{\substack{p \leq \xi \\ (d \mid p) = -1}} \frac{(D \mid p)}{p} = B_3 + \sum_{\substack{p > \xi \\ p \mid d}} \frac{(D \mid p)}{p} + O\left(\frac{1}{\log \xi}\right),$$

and the statement of the lemma follows from this by making $\xi \to \infty$.

Lemma 2. *If $p_1, \cdots, p_s$ are distinct odd primes not dividing N and not all of the Legendre symbols $(N \mid p_1), \cdots, (N \mid p_s)$ are $+1$, there exists an integer v such that*

$$\prod_{i=1}^{s} (N - v^2 \mid p_i) = -\prod_{i=1}^{s} (N \mid p_i).$$

Proof. Suppose that $(N \mid p_1) = -1$. If

$$R_1, R_2, \cdots, R_t \qquad (t = \tfrac{1}{2}(p_1 - 1))$$

are the quadratic residues modulo p_1, at least one of the integers

$$N - R_1, N - R_2, \cdots, N - R_t$$

must be a residue. For, if they were all non-residues, there would be too many non-residues, because N is also a non-residue. Suppose then that $N - R_\alpha$ is a residue. We then choose

$$v^2 \equiv R_\alpha \pmod{p_1},$$

$$v \equiv 0 \pmod{p_i} \qquad (i = 2, \cdots, s).$$

Thus $(N - v^2 \mid p_1) = -(N \mid p_1)$ and $(N - v^2 \mid p_i) = (N \mid p_i)$ $(i = 2, \cdots, s)$.

Lemma 3. *If $p_1, \cdots, p_s$ are distinct odd primes not dividing N and all the Legendre symbols $(N \mid p_1), \cdots, (N \mid p_s)$ are $+1$, there exists an integer v such that $N - v^2 \equiv 0 \pmod{H^2}$ and*

$$\prod_{i=1}^{s} \left(\frac{N - v^2}{H^2} \,\middle|\, p_i\right) = -\prod_{i=1}^{s} (N \mid p_i),$$

where $H = p_1 p_2 \cdots p_s$.

Proof. Let $H = p_1 p_2 \cdots p_s$. There exist solutions of each of the congruences[7]

$$N - v_1^2 \equiv QH^2 \pmod{p_1^3},$$
$$N - v_i^2 \equiv H^2 \pmod{p_i^3} \qquad (i = 2, \cdots, s),$$

where $(Q \mid p_1) = -1$. Then we choose $v \equiv v_i \pmod{p_i^3}$ $(i = 1, 2, \cdots, s)$, so that

$$N - v^2 \equiv 0 \pmod{H^2}$$

and

$$\left(\frac{N - v^2}{H^2}\middle| p_1\right) = -(N \mid p_1), \qquad \left(\frac{N - v^2}{H^2}\middle| p_i\right) = (N \mid p_i) \qquad (i = 2, \cdots, s).$$

3. **The sieve method.** In the paper to which reference was made in §1, Estermann adopts the following notation which we shall also use in this section. The letter q with or without a subscript will always denote a quadratfrei integer, and $\lambda(q)$ will denote the number of prime factors of q. The function $g(n)$ is defined for all positive integers n as follows:

$$g(n) = \begin{cases} 1, & n = 1, \\ 0, & n > 1. \end{cases}$$

A function $w(q)$ is called multiplicative if

$$w(q_1 q_2) = w(q_1)w(q_2).$$

In addition, the usual notation $a \mid b$ for "a divides b", and (a, b) for the greatest common divisor of a and b will be employed.

Instead of the recurrence relations of earlier methods Estermann makes use of the following

LEMMA 4 (Estermann,[4] Hilfsatz 1). *Let*

$$b_0, b_1, \cdots, b_t$$

be a sequence of positive integers such that

$$b_n \mid b_{n-1} \qquad (1 \leq n \leq t).$$

Then

$$g(b_0) \geq \sum_q (-1)^{\lambda(q)},$$

where the summation for q is over all quadratfrei integers for which

$$q \mid b_0, \qquad \lambda\left(\frac{q}{(q, b_n)}\right) \leq 2n - 1 \qquad (1 \leq n \leq t). \tag{3.1}$$

The numerical constants in Estermann's second lemma need to be improved a little for the problem we are considering. By making slightly more precise approximations we prove the lemma in the following modified form.

[7] Dickson, loc. cit., Theorem 17.

LEMMA 5 (Estermann,[4] Hilfsatz 2). *Let $w(q)$ be a multiplicative function such that*

$$w(1) = 1; \qquad 0 \leqq w(q) < 1,\ q > 1;$$

let $a_0, a_1, \cdots, a_t$ be any sequence of positive integers for which

$$\begin{aligned} &a_n \mid a_{n-1} \qquad (1 \leqq n \leqq t), \\ &a_t = 1. \end{aligned} \tag{3.11}$$

Further, suppose that the products

$$\pi_n = \prod (1 - w(p)) \qquad \left(p \,\middle|\, \frac{a_{n-1}}{a_n}\right)$$

satisfy the inequalities

$$\pi_1 > \beta h_0^{-1},$$

$$\pi_n > h_0^{-1} \qquad (2 \leqq n \leqq t),$$

where $\beta = 0.66867$, $h_0 = 1.67$. Then

$$\sum_q (-1)^{\lambda(q)} w(q) \geqq \frac{1}{15{,}000} \prod_{m=2}^{t} \pi_m, \tag{3.12}$$

where the summation for q is over all quadratfrei integers such that

$$q \mid a_0, \qquad \lambda\left(\frac{q}{(q, a_n)}\right) \leqq 2n - 1 \qquad (1 \leqq n \leqq t). \tag{3.13}$$

Proof. The notation used in the proof of this lemma is necessarily rather involved, and we begin by stating what is required. Let

$$E_s = \sum_q (-1)^{\lambda(q)} w(q) \qquad (s = 0, 1, \cdots, t),$$

where the summation for q is over all quadratfrei integers satisfying the relations (3.13) and in addition such that

$$q \,\Big|\, \frac{a_0}{a_s}; \tag{3.14}$$

let

$$E_s^{(r)} = \sum_q (-1)^{\lambda(q)} w(q),$$

where the summation for q is over all quadratfrei integers satisfying the relations (3.13), (3.14), and in addition such that

$$\lambda(q) = r; \tag{3.15}$$

let

$$S_n^{(r)} = \sum_q w(q) \qquad \left(q \,\middle|\, \frac{a_{n-1}}{a_n},\ \lambda(q) = r\right), \tag{3.16}$$

$$\Phi_n = \sum_{\substack{k \\ k+l=2n,\, l \geq 3}} \sum_{l} E_{n-1}^{(k)} S_n^{(l)}; \tag{3.17}$$

write $\tau = \log h_0$, $\rho = -\tau^{-1} \log \beta$.

Estermann then establishes the following results.[8]

$$E_n^{(r)} = \sum_{k+l=r} \sum E_{n-1}^{(k)} S_n^{(l)} \qquad (r < 2n \leqq 2t); \tag{3.21}$$

$$E_n^{(r)} = 0 \qquad (r \geqq 2n,\ n \leqq t); \tag{3.22}$$

$$S_n^{(l)} \leqq \frac{\tau^l}{l!} \qquad (2 \leqq n \leqq t); \tag{3.23}$$

$$S_1^{(1)} < \tau(1 + \rho); \tag{3.24}$$

$$E_t > \left\{E_1 - \sum_{n=2}^{t} h_0^{n-1} \Phi_n\right\} \prod_{m=2}^{t} \pi_m. \tag{3.25}$$

Since $a_t = 1$, the condition (3.14) is included in (3.13) when $s = t$. Hence E_t is the left member of the inequality (3.12) which we have to establish. In view of (3.25) it is clearly sufficient to show that

$$\left\{E_1 - \sum_{n=2}^{t} h_0^{n-1} \Phi_n\right\} > \frac{1}{15{,}000}. \tag{3.3}$$

Estermann finds a bound for Φ_n which holds for $3 \leqq n \leqq t$ and then computes Φ_2 directly. What we shall do is to find a bound for Φ_n which holds for $6 \leqq n \leqq t$ and then compute Φ_2, Φ_3, Φ_4 and Φ_5 directly.

Since Φ_n depends on $E_{n-1}^{(k)}$ and $S_n^{(l)}$ and since (3.23), (3.24) provide us with a bound for $S_n^{(l)}$, our first step is to approximate to $E_{n-1}^{(k)}$. By (3.22) $E_{n-1}^{(k)} = 0$ if $k \geqq 2n - 2$, so that we may assume $k < 2n - 2$. From (3.21), (3.16), the definition of $w(q)$, and (3.24), we obtain

$$E_0^{(0)} = 1, \qquad E_0^{(1)} = 0;$$

$$E_1^{(0)} = 1, \qquad E_1^{(1)} = S_1^{(1)} < \tau(1 + \rho).$$

If now $A_n^{(r)}$ and $B_n^{(r)}$ are defined by the following equations:

$$\begin{aligned} A_1^{(0)} &= 1, \qquad B_1^{(0)} = 0; \\ A_1^{(1)} &= 1, \qquad B_1^{(1)} = 1; \end{aligned} \tag{3.41}$$

[8] He uses $\beta = \frac{11}{8}$, but there is no change required in the proof for our value of β.

$$
A_n^{(r)} = \sum_{k=0}^{r} \binom{r}{k} A_{n-1}^{(k)},
$$

(3.41)
$$
B_n^{(r)} = \sum_{k=0}^{r} \binom{r}{k} B_{n-1}^{(k)}, \qquad (r < 2n);
$$

$$
A_n^{(r)} = B_n^{(r)} = 0 \qquad (r \geqq 2n),
$$

it is easy to prove by induction, if we use (3.21) and (3.23), that

(3.42)
$$
r!\tau^{-r}E_n^{(r)} < A_n^{(r)} + \rho B_n^{(r)}.
$$

From the definition of Φ_n, (3.42), and (3.23) it follows that

$$
\begin{aligned}
\Phi_n &\leqq \sum_{k} \sum_{\substack{l \\ k+l=2n,\, l \geqq 3}} \frac{\tau^k}{k!} \{A_{n-1}^{(k)} + \rho B_{n-1}^{(k)}\} \frac{\tau^l}{l!} \\
&= \frac{\tau^{2n}}{(2n)!} \sum_{k=0}^{2n-3} \binom{2n}{k} \{A_{n-1}^{(k)} + \rho B_{n-1}^{(k)}\}.
\end{aligned}
$$

The values of $A_{n-1}^{(k)}$ and $B_{n-1}^{(k)}$ may be computed from (3.41), and we find that

$$
\Phi_2 \leqq \frac{\tau^4}{4!} (5 + 4\rho),
$$

$$
\Phi_3 \leqq \frac{\tau^6}{6!} (138 + 96\rho),
$$

$$
\Phi_4 \leqq \frac{\tau^8}{8!} (8915 + 5832\rho),
$$

$$
\Phi_5 \leqq \frac{\tau^{10}}{10!} (1037240 + 655360\rho).
$$

Then from the definition of h_0, τ, β, and ρ we obtain $\rho < 0.7845$, and thus we have

(3.5)
$$
\begin{aligned}
h_0\Phi_2 &< 0.03922, \\
h_0^2\Phi_3 &< 0.01507, \\
h_0^3\Phi_4 &< 0.00748, \\
h_0^4\Phi_5 &< 0.00420.
\end{aligned}
$$

We turn now to the approximation for Φ_n when $n \geqq 6$. Estermann shows that if $2^k\tau^{-k}E_{n-1}^{(k)} \leqq \alpha e^{2n-6}$ for every k, then $2^r\tau^{-r}E_n^{(r)} \leqq \alpha e^{2n-4}$ for every r. Here α is independent of n, k, and r. Hence, if

(3.61)
$$
2^k\tau^{-k}E_5^{(k)} \leqq \alpha e^6
$$

for every k, then

$$
2^r\tau^{-r}E_n^{(r)} \leqq \alpha e^{2n-4}
$$

for every r and every $n \geqq 5$. It then follows from (3.17) and (3.23) that

$$\Phi_n < \alpha e^{2n-6}(\tfrac{1}{2}\tau)^{2n}(e^2 - 5)$$

for every $n \geqq 6$. Then

$$\sum_{n \geqq 6} h_0^{n-1}\Phi_n < \alpha h_0^5(\tfrac{1}{2}\tau)^{12}(e^2 - 5)(1 - h_0 e^2 \tfrac{1}{4}\tau^2)^{-1}. \tag{3.62}$$

Our problem then is to find a suitable α which satisfies (3.61). Because of the inequality (3.41) the largest of the numbers

$$\frac{2^k}{k!}(A_5^{(k)} + \rho B_5^{(k)}) \qquad (k = 0, 1, \cdots, 9)$$

will serve as a value for α. Numerical computation shows that $k = 7$ yields a maximum and this gives

$$\alpha < 1351.6.$$

Then from (3.62) we have

$$\sum_{n \geqq 6} h_0^{n-1}\Phi_n < 0.01851. \tag{3.63}$$

Next, from the definitions of E_s and $E_s^{(r)}$ we obtain

$$\begin{aligned} E_1 &= E_1^{(0)} - E_1^{(1)}, \\ E_1 &> 1 - \tau(1 + \rho) > 0.08455. \end{aligned} \tag{3.7}$$

Finally, from (3.5), (3.63), and (3.7) it follows that

$$E_1 - \sum_{n=2}^{t} h_0^{n-1}\Phi_n > 0.00007 > \frac{1}{15{,}000},$$

and, in view of (3.3), this completes the proof of the lemma.

4. **Application of the sieve method.** Our object in this section is to apply the results of §3 to the particular function $w(q)$ and the particular sequence of integers $a_0, a_1, \cdots, a_t$ defined as follows. Let

$$w(q) = \prod_{\substack{p \mid q \\ (d \mid p) = -1}} \left\{\frac{1 + (D \mid p)}{p}\right\},$$

where d and D are any integers for which the Kronecker symbols $(d \mid m)$ and $(D \mid m)$ are defined and such that $dD < 0$. It is easily seen that this function satisfies the conditions of Lemma 5. Let N and c be two integers, where c is to be defined later and

$$N > c^{6h^{20}\beta^{-1}}, \qquad h = \sqrt[4]{2.7888}. \tag{4.11}$$

Then, the integers $a_0, a_1, \cdots, a_t$ are given by

$$a_0 = \prod p, \qquad a_n = \prod p, \tag{4.12}$$

where the ranges of the p in the two products are, respectively,

$$c < p \leqq N^{\frac{1}{2}}, \qquad (d \mid p) = -1,$$

$$c < p \leqq N^{\frac{1}{2}\beta^2 h^{-4n}}, \qquad (d \mid p) = -1.$$

The integer t is defined to be the least integer for which $a_t = 1$. We must, of course, show that t exists and that $t \geqq 6$. To do this we make use of Lemma 1 and the well-known result that

$$\lim_{\xi \to \infty} \left\{ \sum_{\substack{p \leqq \xi \\ (d \mid p) = -1}} \frac{1}{p} - \frac{1}{2} \log \log \xi \right\}$$

exists. Thus

$$\lim_{\xi \to \infty} \left\{ \sum_{\substack{p \leqq \xi \\ (d \mid p) = -1}} \left(\frac{1 + (D \mid p)}{p} \right) - \frac{1}{2} \log \log \xi \right\}$$

exists and hence so does

$$\lim_{\xi \to \infty} \left\{ \sum_{\substack{p \leqq \xi \\ (d \mid p) = -1}} \log \left\{ 1 - \frac{1 + (D \mid p)}{p} \right\} + \frac{1}{2} \log \log \xi \right\}.$$

Since $h^2 < h_0$, this means that there is a constant $c \geqq 3$ such that for $\xi \geqq c$ we have

$$\left| \sum_{\substack{c < p \leqq \xi \\ (d \mid p) = -1}} \log \left\{ 1 - \frac{1 + (D \mid p)}{p} \right\} + \frac{1}{2} \log \log \xi - \frac{1}{2} \log \log c \right| \tag{4.2}$$
$$< \tfrac{1}{2} \log h_0 - \log h.$$

If there were no prime p such that $(d \mid p) = -1$ and $c < p \leqq c^2$, it would follow from (4.2) with $\xi = c^2$ that

$$\tfrac{1}{2} \log \log c^2 - \tfrac{1}{2} \log \log c < \tfrac{1}{2} \log h_0 - \log h,$$

$$\tfrac{1}{2} \log 2 < \tfrac{1}{2} \log h_0 - \log h.$$

This is a contradiction since

$$\frac{h_0}{h^2} = \frac{1.67}{\sqrt{2.7888}} < 2,$$

$$\tfrac{1}{2} \log h_0 - \log h < \tfrac{1}{2} \log 2.$$

Then there is at least one prime p such that $(d \mid p) = -1$ between c and c^2, and this means that

$$N^{\frac{1}{2}\beta^2 h^{-4t}} < c^2.$$

From this inequality and (4.11) we obtain $t > 5$.

Furthermore

$$w(p) = \frac{1 + (D \mid p)}{p} \tag{4.3}$$

when p is a prime such that $(d \mid p) = -1$. By (4.3), (4.12), and (4.2) we have

$$\begin{aligned}
\log \pi_1 &= \log \prod (1 - w(p)) = \log \prod \left(1 - \frac{1 + (D \mid p)}{p}\right) && \left(p \,\middle|\, \frac{a_0}{a_1}\right) \\
&> -\tfrac{1}{2} \log \log N^{\frac14} + \tfrac{1}{2} \log \log N^{\frac14 \beta^{2\lambda - 4}} - \log h_0 + 2 \log h \\
&= \log \beta - \log h_0; \\
\log \pi_n &= \log \prod (1 - w(p)) && \left(p \,\middle|\, \frac{a_{n-1}}{a_n}\right) \\
&> -\tfrac{1}{2} \log \log N^{\frac14 \beta^{2\lambda - 4n}} + \tfrac{1}{2} \log \log N^{\frac14 \beta^{2\lambda - 4n + 4}} - \log h_0 + 2 \log h \\
&= - \log h_0 .
\end{aligned}$$

Hence all the conditions of Lemma 5 are satisfied, and thus

$$\begin{aligned}
\sum_q (-1)^{\lambda(q)} w(q) &> \frac{1}{15{,}000} \prod_{m=2}^{t} \pi_m \\
&= \frac{1}{15{,}000} \prod_{p \mid a_1} (1 - w(p)).
\end{aligned}$$

5. **Proof of the main theorem.** We shall show first of all that there exists an integer w such that every integer N is represented in the form

$$N = w^2 + H^2 P \tag{5.11}$$

with the exception of the cases noted in §1, when N is represented in the form

$$N = w^2 + 2H^2 P. \tag{5.12}$$

In (5.11) and (5.12) every prime dividing H also divides d and P is an integer such that $(d \mid P) = -1$. We distinguish three cases.

I. Suppose that $(d \mid N) = -1$. Then (5.11) is satisfied with $w = 0$, $H = 1$, $P = N$.

II. Suppose that $(d \mid N) = 1$. If $d = -2^b S^2 h$ we have (see footnote 5)

$$(d \mid N) = \begin{cases} (N \mid h), & b \text{ even}, \quad h \equiv 3 \pmod 4, \\ (2 \mid N)(N \mid h), & b \text{ odd}, \quad h \equiv 3 \pmod 4, \\ (-1 \mid N)(N \mid h), & b \text{ even}, \quad h \equiv 1 \pmod 4, \\ (-2 \mid N)(N \mid h), & b \text{ odd}, \quad h \equiv 1 \pmod 4. \end{cases}$$

Let $p_1, p_2, \cdots, p_n$ be the distinct odd primes dividing h. If not all the symbols $(N \mid p_i)$ are $+1$, Lemma 2 shows that there exists an integer v such that

$$\prod_{i=1}^{n} (N - v^2 \mid p_i) = -\prod_{i=1}^{n} (N \mid p_i).$$

We now choose w so that $w \equiv 0 \pmod 4$, $w \equiv v$ or $0 \pmod h$. Then in all cases we have $(d \mid N - w^2) = -(d \mid N)$ and (5.11) is satisfied with $H = 1$.

If all the symbols $(N \mid p_i)$ are $+1$, we use Lemma 3 and conclude that there exist integers v and H such that

$$N - v^2 \equiv 0 \pmod{H^2},$$

$$\left(d \,\middle|\, \frac{N - v^2}{H^2}\right) = -(d \mid N) = -1,$$

and again (5.11) can be satisfied.

This proof does not apply if $h = 1$, for then

$$(d \mid N) = \begin{cases} (-1 \mid N), & b \text{ even}, \\ (-2 \mid N), & b \text{ odd}. \end{cases}$$

However, in the respective cases we have

$$\left(-1 \,\middle|\, \frac{N - v^2}{4}\right) = -1, \qquad N \equiv 5 \pmod 8,$$

$$\left(-1 \,\middle|\, \frac{N - v^2}{16}\right) = -1, \qquad N \equiv 1 \pmod 8,$$

$$(-2 \mid N - 4) = -1, \qquad N \equiv 1 \text{ or } 3 \pmod 8.$$

III. Suppose that $(d \mid N) = 0$. Let $h = h_1 h_2$, where $h_1 \mid N$ and no prime which divides h_2 also divides N. Let $p_1, p_2, \cdots, p_n$ be the distinct odd primes dividing h_2. If not all the symbols $(N \mid p_i)$ are $+1$, there is an integer v such that

$$\prod_{i=1}^{n} (N - v^2 \mid p_i) = -\prod_{i=1}^{n} (N \mid p_i).$$

If we choose $w \equiv 0$ or $1 \pmod{4S^2}$, $w \equiv 1 \pmod{h_1}$, $w \equiv v$ or $0 \pmod{h_2}$, we can make

$$(d \mid N - w^2) = -1.$$

Similarly, if all the symbols $(N \mid p_i)$ are $+1$, we can make

$$\left(d \,\middle|\, \frac{N - w^2}{H^2}\right) = -1.$$

The proof fails if $h_2 = 1$, but by taking H equal to an appropriate power of 2, we can still make

$$\left(d \,\middle|\, \frac{N - w^2}{H^2}\right) = -1$$

with the exceptions stated in the theorem. In these exceptional cases we can show that

$$\left(d \,\middle|\, \frac{N - w^2}{2H^2}\right) = -1.$$

This completes the proof that N can be represented in one of the forms (5.11), (5.12).

6. **Proof of the main theorem** (continued). In this section we shall show that for every sufficiently large integer N there exist integers u, H with the following properties:

(6.11) $$u < N^{1/2};$$

(6.12) every prime p satisfying $(d \mid p) = -1$ which divides $N - u^2$ is greater than $N^{1/3}$;

(6.13) every prime dividing H also divides d;

(6.14) either $N - u^2 \equiv 0 \pmod{H^2}$ or $N - u^2 \equiv 0 \pmod{2H^2}$;

(6.15) either $\left(d \,\middle|\, \frac{N - u^2}{H^2}\right) = -1$ or $\left(d \,\middle|\, \frac{N - u^2}{2H^2}\right) = -1$.

It follows from (6.14), (6.15) that

(6.2) $$N = u^2 + H^2Q \quad \text{or} \quad N = u^2 + 2H^2Q,$$

where in each case $(d \mid Q) = -1$.

Now suppose that the product of primes $p_1p_2 \cdots p_r$, not necessarily distinct, are all the odd primes satisfying $(d \mid p) = -1$ which divide $N - u^2$. Their product is greater than $N^{r/3}$ by (6.12) and hence $r \leq 2$. The product must divide Q, and since $(d \mid Q) = -1$, r must be odd. This means that $r = 1$. Thus $Q = pM$, where p and M have the required properties of the theorem.

We proceed to the proof of (6.11)-(6.15). Let

$$f = \prod_{\substack{p^3 \leq N \\ (d \mid p) = -1}} p, \qquad g = \prod_{\substack{p \leq c \\ (d \mid p) = -1}} p,$$

where c was defined in §4. Then from (4.12) we have

$$f = ga_0, \qquad (g, a_0) = 1.$$

There exists an integer x such that no prime p of the product g divides $N - x^2$. For if p is a representative prime, we choose $x \equiv 1$ or $0 \pmod{p}$ according as $p \mid N$ or $p \nmid N$. Since $(dH^2, g) = 1$, we can find an integer y such that

$$y \equiv x \pmod{g},$$
$$y \equiv w \pmod{dH^2}.$$

Then y is determined modulo gdH^2, no prime of the product g divides $N - y^2$, and

$$\left(d \left| \frac{N - y^2}{H^2}\right.\right) = -1.$$

Now consider the sum

$$S = \sum_{u^2 \leqq N} 1,$$

where the summation is over all integers $u \leqq N^{1/2}$ satisfying $(N - u^2, f) = 1$ and the conditions (6.14), (6.15). We have to prove that $S > 0$, for then there is at least one value of u such that *no* prime p satisfying $(d \mid p) = -1$ and less than $N^{1/2}$ divides $N - u^2$. Clearly

$$S \geqq \sum_{u^2 \leqq N}' 1, \tag{6.1}$$

where the summation is over all integers $u \leqq N^{1/2}$ such that $(N - u^2, a_0) = 1$, $u \equiv y \pmod{m}$, where $m = gdH^2$. Using the function $g(n)$ defined in §3, we may write (6.1) in the form

$$S \geqq \sum_{\substack{u^2 \leqq N \\ u \equiv y \pmod{m}}} g\{(N - u^2, a_0)\}.$$

Let $b_n = (N - u^2, a_n)$. Since $a_n \mid a_{n-1}$, we have $b_n \mid b_{n-1}$. Then from Lemma 4 we obtain

$$g\{(N - u^2, a_0)\} \geqq \sum_q (-1)^{\lambda(q)},$$

where the summation for q is given by (3.1) with b_n replaced by $(N - u^2, a_n)$.

The conditions (3.1) with $b_n = (N - u^2, a_n)$ are equivalent to the conditions (3.13) and the additional restriction

$$q \mid (N - u^2). \tag{6.2}$$

Hence

$$\begin{aligned} S &\geqq \sum_{\substack{u^2 \leqq N \\ u \equiv y \pmod{m}}} \sum_{q \mid (N - u^2)} (-1)^{\lambda(q)} \\ &= \sum_q (-1)^{\lambda(q)} \sum_{\substack{u^2 \leqq N \\ u \equiv y \pmod{m},\ u^2 \equiv N \pmod{q}}} 1, \end{aligned} \tag{6.3}$$

where the summation for q is now given by (3.13).

We now evaluate the inner sum. Since q is quadratfrei and odd, the number of solutions of $u^2 \equiv N \pmod{q}$ in any residue system mod q is

$$\prod_{p \mid q} \{1 + (N \mid p)\}. \tag{6.4}$$

Furthermore $(q, m) = 1$ when $q \mid a_0$, because a_0 involves only primes $p > c$. Thus the number of solutions of the simultaneous congruences

$$u^2 \equiv N \pmod{q}, \qquad u \equiv y \pmod{m}$$

in any residue system mod qm is also (6.4).

In the function $w(q)$ in §3 let $D = 4N$. Thus

$$\begin{aligned}\prod_{p \mid q} \{1 + (N \mid p)\} &= \prod_{p \mid q} \{1 + (D \mid p)\} \\ &= qw(q).\end{aligned}$$

We therefore have

$$\sum_{\substack{u^2 \leqq N \\ u^2 \equiv N \pmod{q},\ u \equiv y \pmod{m}}} 1 - \left[\frac{N^{\frac{1}{2}}}{qm}\right] qw(q) \leqq \prod_{p \mid q} \{1 + (D \mid p)\} \leqq 2^{\lambda(q)},$$

$$\sum_{\substack{u^2 \leqq N \\ u^2 \equiv N \pmod{q},\ u \equiv y \pmod{m}}} 1 - \left\{\left[\frac{N^{\frac{1}{2}}}{qm}\right] + 1\right\} qw(q) \geqq -2^{\lambda(q)}.$$

From this we obtain

$$\left| \sum_{\substack{u^2 \leqq N \\ u^2 \equiv N \pmod{q},\ u \equiv y \pmod{m}}} 1 - \frac{N^{\frac{1}{2}}}{qm} qw(q) \right| \leqq 2^{\lambda(q)}.$$

Then (6.3) becomes

$$S \geqq \frac{N^{\frac{1}{2}}}{m} \sum_{q} (-1)^{\lambda(q)} w(q) - \sum_{q} 2^{\lambda(q)}$$

and, using (4.4) we have

$$S \geqq \frac{N^{\frac{1}{2}}}{15{,}000\, m} \prod_{p \mid a_1} \{1 - w(p)\} - \sum_{q} 2^{\lambda(q)}. \tag{6.5}$$

As in §4 we obtain

$$\begin{aligned}\log \prod_{p \mid a_1} (1 - w(p)) &= \sum \log \left\{1 - \frac{1 + (D \mid p)}{p}\right\} \\ &> -\tfrac{1}{2} \log\log N^{\frac{1}{2}\delta^2 h^{-4}} + \tfrac{1}{2} \log\log c - \tfrac{1}{2} \log h_0 + \log h \\ &> -\tfrac{1}{2} \log\log N,\end{aligned}$$

where the range for p in the summation is

$$c < p \leqq N^{\frac{1}{2}\delta^2 h^{-4}}, \qquad (d \mid p) = -1.$$

Hence

$$\prod_{p \mid a_1} (1 - w(p)) > (\log N)^{-\frac{1}{2}}. \tag{6.6}$$

Estermann has shown that

$$\sum_q 2^{\lambda(q)} < 4\lambda(a_0) \prod_{n=1}^{t-1} \{2\lambda(a_n)\}^2,$$

and in our case

$$2\lambda(a_0) < N^{\frac{1}{3}},$$

$$2\lambda(a_n) < N^{\frac{1}{3}\beta^2 h^{-4n}}.$$

This proves that

$$\sum_q 2^{\lambda(q)} < 2N^{\frac{1}{3}(1+2\beta^2(h^4-1)^{-1})}. \tag{6.7}$$

Then from (6.5), (6.6), (6.7) we obtain

$$S \geqq \frac{1}{15{,}000\, m}\left(\frac{N}{\log N}\right)^{\frac{1}{2}} - 2N^{\frac{1}{3}(1+2\beta^2(h^4-1)^{-1})}.$$

Finally, comparing the exponents of N we see that

$$\tfrac{1}{3}(1 + 2\beta^2(h^4 - 1)^{-1}) = \frac{1}{3}\left(1 + 2\,\frac{(0.66867)^2}{1.7888}\right) < \tfrac{1}{2}.$$

This shows that $S > 0$ when N is sufficiently large. The proof of the theorem is then complete.

University of Saskatchewan.

rem

CONTENTS